JOURNAL OF APPLIED PHYSICS

JOURNAL
OF
APPLIED PHYSICS

Vol. 21, No. 3 MARCH, 1950

In This Issue

Page

Editorial: British Journal of Applied Physics 187

Microwave Angle Separation on a Two and One-Half Mile Overwater Path A. W. Straiton and A. H. LaGrone 188

Immobility Phenomena and Reverse Driving Phenomena of the Electric Arc Sakae Yamamura 193

Measurements of Mechanical Properties of Polyisobutylene at Audio-frequencies by a Twin Transducer Robert S. Marvin, Edwin R. Fitzgerald, and John D. Ferry 197

Dielectric Strength of Gaseous Fluorocarbons W. A. Wilson, J. H. Simons, and T. J. Brice 203

Complex Stressing of Polyethylene I. L. Hopkins, W. O. Baker, and J. B. Howard 206

D. C. Characteristics of Silicon and Germanium Point Contact Crystal Rectifiers. Part I. Experimental H. J. Yearian 214

Electric Breakdown in CO_2 from Low Pressures to the Liquid State D. R. Young 222

Method for the Numerical Calculation of Hydrodynamic Shocks J. Von Neumann and R. D. Richtmyer 232

Some Non-Linear Systems Permitting Simple Harmonic Motion John E. Brock 238

Analogy between Hydraulic Jumps in Liquids and Shock Waves in Gases F. R. Gilmore, M. S. Plesset, and H. E. Crossley, Jr. 243

Optical Method for Measuring the Stress in Glass Bulbs W. T. Read 250

Letters to the Editor:

Note on the Frictionless Bearing for Small Angular Deflections J. A. Haringx 258
Reply to J. A. Haringx Helmut Schlitt 258
Comments on Creep and Damping Properties of Polystyrene William N. Findley 258
On the Growth of Metallic Crystals from the Vapor Phase George R. Keepin, Jr. 260
Thermionic Emitting Properties of Thoria-Rhenium G. A. Espersen 261
Activation Energy for Copper Self-Diffusion John H. Hollomon 261
Reply to "Activation Energy for Copper Self-Diffusion" J. H. Dedrick and A. Gerds 261
Study of Crack Propagation Using High Speed Motion Pictures Herbert I. Fusfeld and Josephine Carr Feder 261

Index to Advertisers xxviii

BOARD OF EDITORS
Elmer Hutchisson,
Editor

Associate Editors
Howard H. Aiken
Otto Beeck
James Hillier
W. James Lyons
Philip Morrison
Eric Reissner
Raymond J. Seeger
Frederick Seitz
Charles S. Smith

Publication of the
American Institute of Physics

OFFICERS
George R. Harrison,
Chairman
George B. Pegram,
Treasurer
John T. Tate,
Adviser on Publications

GOVERNING BOARD
Luis W. Alvarez
J. W. Beams
Wallace R. Brode
E. U. Condon
W. F. Fair, Jr.
Harvey Fletcher
George R. Harrison
W. V. Houston
Paul Kirkpatrick
Paul E. Klopsteg
F. W. Loomis
Philip M. Morse
J. R. Oppenheimer
George B. Pegram
Harold K. Schilling
John T. Tate
Wallace Waterfall
M. W. White

Consultant on Printing
Melvin Loos

ADMINISTRATIVE STAFF
Henry A. Barton,
Director
Wallace Waterfall,
Executive Secretary
Theodore Vorburger,
Assistant Treasurer and Comptroller
Ruth F. Bryans,
Publication Manager
Betty Hofstadter,
Assistant Secretary
Eileen B. Neuberger,
Circulation Manager

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members and associate members of the American Institute of Physics	$7.00	$ 8.00
To all others	9.00	10.00

Back Number Prices

Complete set: Vol. 1, July, 1931–Vol. 19, 1949—$148.00
Yearly back number rate: $10.00
Single copies: $1.00 each.

Changes of address should be sent to the Circulation Manager, American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 21, Number 3 March, 1950

British Journal of Applied Physics

IT is a pleasure to welcome a new journal to the fold of physical literature. In January of this year there appeared Volume One, Number One, of the *British Journal of Applied Physics*. This new journal is published by the Institute of Physics in Great Britain under the editorship of Dr. H. R. Lang, the Secretary of the Institute. In a foreword in this first issue, Dr. T. C. Toy, President of the Institute of Physics, mentions the interest which the Institute has had in applied physics since 1918. The first fruit of this interest was the inauguration in 1923 of a series of monographs and lectures under the general title of "Physics in Industry." This was followed by the establishment of the Institute's *Journal of Scientific Instruments*. Now the parts of this journal related particularly to the new applications of physics and new developments have been separated and published together as the *British Journal of Applied Physics*.

The first issue of this new journal sets a high standard indeed. The two leading articles are of a general nature, one entitled "A Scientific Education" by Sir Philip Morris, and the other on "Some Chemical and Physical Properties of Rubber" by J. Moore. There then follow five original contributions on the Development and Deionization Time of Heavy-Current A.C. Arcs, Measurement of Opacity and Reflectivity for Printing Papers, Measurement of the Coefficient of Internal Friction of Solid Rods by a Resonance Method, The Application of Multiple Factor Analysis to Industrial Test Data, and a Note on the Interpretation of Multiple Factor Analysis. In addition, this issue reviews several new books of particular interest to applied physicists.

The new journal is printed on a high quality coated stock which allows excellent reproduction of halftones. The page size and general appearance are similar to that of the *Journal of Scientific Instruments*. A small number of advertisements are also included.

There is an interesting and significant parallel in the development of applied physics in Great Britain and in the United States. As mentioned, the Institute of Physics was established in Great Britain in 1918. In 1931 the several national organizations devoted to physics and its applications in this country established the American Institute of Physics. The first publication to come under the direct sponsorship of the American Institute was the *Review of Scientific Instruments*. An Advisory Council on Applied Physics was established, and the theme of the large conference celebrating the fifth anniversary of the founding of the American Institute was "Physics in Industry." Shortly thereafter, in 1937, the *Journal of Applied Physics* was launched. Now the parallel is completed with the publication of the *British Journal of Applied Physics*. Such a close parallel is indeed an indication of the real place that exists in each country for a journal devoted to the interests of the applied physicist. Significantly journals of almost the same name and purpose have recently been established in all of the principal countries of Europe.

We want in this brief note to congratulate Dr. Lang on this new venture of the Institute of Physics and join Dr. Toy in requesting "our many friends all over the world to support this journal by contributing to its pages, by subscribing to it, and by encouraging others to do so also."

ELMER HUTCHISSON

Microwave Angle Separation on a Two and One-Half Mile Overwater Path*

A. W. Straiton and A. H. LaGrone
The University of Texas, Austin, Texas
(Received August 29, 1949)

This paper describes the results of radio angle-separation measurements on a 2.50-mile path over Lake Buchanan near Austin, Texas. The measurements were made to test the angle-separation equipment developed at The University of Texas, and to compare these results with measurements obtained by rocking a 20-foot parabolic antenna.

I. INTRODUCTION

THE measurement of the angle-of-arrival of radio waves at near grazing angles is complicated by reflections from the ground. This paper describes measurements made to separate the direct wave from the ground-reflected wave, using a wave-length of 3.2 centimeters.

The measurements made in this report were for two rays arriving at angles separated by less than 0.5 degree. The magnitude of the reflected ray was measured to be 0.94 of the direct ray. Thus, these conditions of two rays of approximately equal magnitude arriving at such small angles of separation made the problem of their separation a very difficult one.

The angles measured by the phase-difference method are not singular, but are ambiguous. For a spacing of 5 feet and a wave-length of 3.2 centimeters, this ambiguity occurs every 1.2 degrees. This difficulty could be resolved by coarser methods of measurement.

Three different methods were used in the angle determinations.

1. The first method requires as data the relative phase and the relative signal strength of three vertically spaced antennas. This method will be referred to throughout this report as the "three-antenna method."

2. The second method requires as data the relative phase between two antennas, the signal strength at the two antennas relative to the maximum signal from a height-gain curve, and the minimum signal from a height-gain curve. This method will be referred to throughout this report as the "two-antenna method."

3. The third method requires as data the record of the signal received by a large antenna as its axis is tilted through an angle large enough to include the two-wave components. This method will be referred to throughout this report as the "amplitude method."

II. THEORY OF MEASUREMENT METHODS

1. Three-Antenna Method

This method assumes two plane or two spherical waves arriving at three vertically spaced antennas. If the resultant phase differences between the three antennas and the resultant relative signal strengths at the three antennas are known, the angles-of-arrival of each of the two-wave components, their relative magnitudes and their relative phase at the center antenna may be determined by calculations and charts. This mathematical relationship has been described elsewhere.[1,2]

Equipment has been built at the Electrical Engineering Research Laboratory of The University of Texas for obtaining these desired data. The basic phase-difference equipment has been described in Report 8,[3] and the operation characteristics of the three-antenna equipment has been described in Report 24.[4]

This method does not require moving the antenna system and hence gives instantaneous data.

2. Two-Antenna Method

The two-antenna method requires a height run over a wide enough range to include at least one signal strength minimum and one maximum. In addition, the phase difference between two points and the signal strength of these points relative to the signal strength maximum must be known.

For the five-foot spacing between antennas used in this work, the angle-of-arrival of the direct wave, θ_D, and the angle-of-arrival of the reflected wave, θ_R, are given by

$$\theta_D = -(1/300)(d\beta + A_z - A_0), \tag{1}$$

$$\theta_R = \theta_D - (1/300)(T_z - T_0). \tag{2}$$

In these equations, $d\beta$ is the phase of the upper antenna signal minus the phase of the lower antenna signal. A_z and T_z are terms which depend on the signal at the upper antenna relative to the maximum signal strength. A_0 and T_0 are the corresponding terms for the lower antenna. If the signal strength is increasing with height at the point in question, A is positive and T is negative. If the signal strength is decreasing with height, A is negative and T is positive.

The derivation of the basic relationships and the

* This work was sponsored by ONR, Contract N5ori-136, Project Order I.

[1] Hamlin, Seay, and Gordon, "New solutions to the problem of vertical angle-of-arrival of radio waves," J. App. Phys. **20**, 248 (1949).

[2] A. H. LaGrone and J. A. Hawkins, "Application of angle-of-arrival to two spherical waves," Memorandum No. 5, Electrical Engineering Research Laboratory, The University of Texas, 1948.

[3] F. E. Brooks, Jr. *et al.*, "Operating manual for phase difference equipment," Report No. 8, Electrical Engineering Research Laboratory, The University of Texas, 1948.

[4] A. W. Straiton *et al.*, "Operational Characteristics of angle separation equipment," Report No. 24, Electrical Engineering Research Laboratory, The University of Texas, 1949.

charts for obtaining A and T are given elsewhere.[5] The minus sign in front of the right-hand expression of Eq. (1) was omitted in reference 5.

3. Amplitude Method

This method involves the use of a large antenna which is tilted so as to determine its scanning pattern in a vertical plane. The antenna used was a 20-foot parabolic antenna with a rectangular face. This antenna was used by the Bell Telephone Laboratories in their early microwave angle-of-arrival measurements,[6] and was subsequently made available to the Electrical Engineering Research Laboratory. It is assumed that maxima in the scanning pattern occur when the tilt is such that the axis of the parabola is pointed in the direction of an incoming wave component.

III. MEASUREMENT SITE

1. Location

The radio path chosen was across Lake Buchanan located about 65 miles from Austin. The length of the path was about two and one-half miles. This length of path was chosen because (1) it made possible a range of angles up to about 0.5 degree; (2) it was long enough to reduce the effect of wave curvature; and (3) it was short enough to make the effect of meteorological changes small.

2. Transmitter

The transmitter was located on a 50-foot tower approximately 20 feet from the water's edge, and had a minimum height of 11 feet above the water. The transmitter assembly used a 2K39 Klystron with a c.w. output of 200 milliwatts.

3. Receivers

The 60-foot receiving tower, with the three receiving antennas mounted on an elevator, was located about 300 feet from the water's edge. The minimum height of the center antenna was 22 feet above the lake level. The vertical spacing of the three-antenna array was five feet. A reference transmitter was moved in front of the receiving antenna and surveyed into position when a reference angle was needed.

The Bell Laboratories' antenna was located near the 60-foot receiving tower and about 250 feet from the water's edge. A synchronous motor was provided for rocking the antenna at a uniform rate. However, in taking the scanning patterns used in this report, the antenna was tilted by hand in order to obtain its position more accurately. The center of the 20-foot parabola was 30 feet above the water.

[5] Straiton, Gordon, and LaGrone, "A method of determining the angle-of-arrival," J. App. Phys. **19**, 524–533 (1948).

[6] W. M. Sharpless, "Measurement of the angle-of-arrival of microwaves," Proc. Inst. Radio Eng. **34**, 837–845 (1946).

4. Meteorological Station

In order to compensate for changes in the meteorological conditions near the ground, a meteorological station was set up near the receiving site. A tilted 60-foot mast imbedded at the water's edge permitted raising a recording psychrometer a short distance out from shore in the direction of the prevailing wind. An anemometer was located near by to make a record of the wind speed and direction.

5. Angles-of-Arrival under Standard Conditions

The angular separation of the direct and reflected rays is a function of the height of the transmitter and the path length for standard meteorological conditions. This angular separation is shown for various transmitter heights in Fig. 1.

The angles-of-arrival of the direct and reflected rays are functions of transmitter and receiver height and path length. These angles-of-arrival as a function of transmitter height are plotted in Fig. 1 for a number of receiver heights with correction applied for the refraction associated with standard meteorological conditions.

IV. RESULTS OF MEASUREMENTS

1. Three-Antenna Method

Measurements were made of the angle-of-arrival by the three-antenna method for 164 cases. These were made for various combinations of transmitter and re-

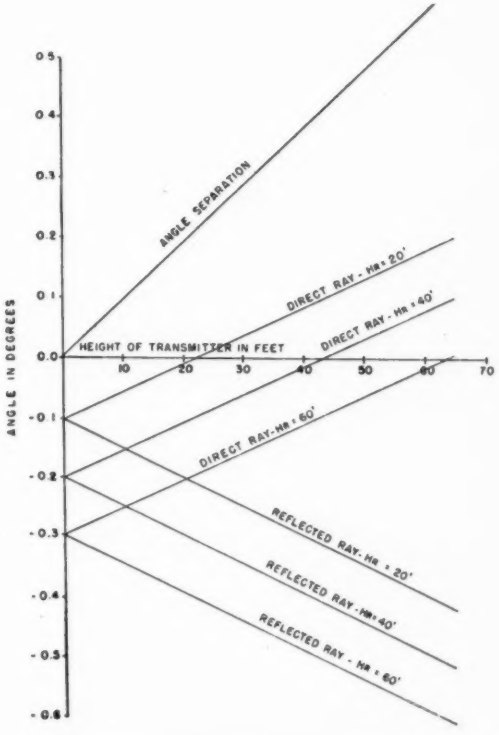

FIG. 1. Angles as a function of transmitter and receiver height.

TABLE I. Errors in angles-of-arrival of the direct and reflected rays.* Three-antenna method.

Transmitter height	Separation angle	Direct ray $.00° < \alpha \leq .02°$	$.02° < \alpha \leq .06°$	$.06° < \alpha \leq .10°$	$.10° < \alpha$	Reflected ray $.00° < \rho \leq .02°$	$.02° < \rho \leq .06°$	$.06° < \rho < .10°$	$.10° < \rho$
11 feet	0.10°	11 (2)	33 (18)	8 (4)	2 (2)	3 (3)	5 (2)	4	42 (30)
21 feet	0.19°	2	1	1	3	2	1	3	1
33 feet	0.29°	14 (7)	18	4	18 (16)	11 (2)	15 (3)	8 (3)	20 (15)
43 feet	0.38°	0	1 (1)	0	1 (1)	1 (1)	0	0	1 (1)
53 feet	0.47°	8 (1)	21 (5)	3 (2)	15 (14)	6 (2)	22 (4)	7	12 (11)
		21.3% (35.2%)	45.1% (47.3%)	9.8% (11.0%)	23.8% (5.6%)	14.0% (18.3%)	26.2% (41.5%)	13.4% (17.1%)	46.3% (23.2%)

* Numbers in parentheses are the number of cases which were rejected. Percentages in parentheses are based on the number of accepted cases.

ceiver heights. More cases were considered with the receiver at 30 and 35 feet than at any other height, since this was approximately the height of the center of the 20-foot antenna. More cases were considered with the transmitter at 11 feet than any other height, since this presented the smallest angular separation of the direct and reflected rays.

From the phase differences and signal-strength ratios between the three antennas, the angles-of-arrival of the direct and reflected rays were calculated by the method of reference 1.

The modified index-of-refraction profile was determined from meteorological measurements made simultaneously with the radio measurements. Since the

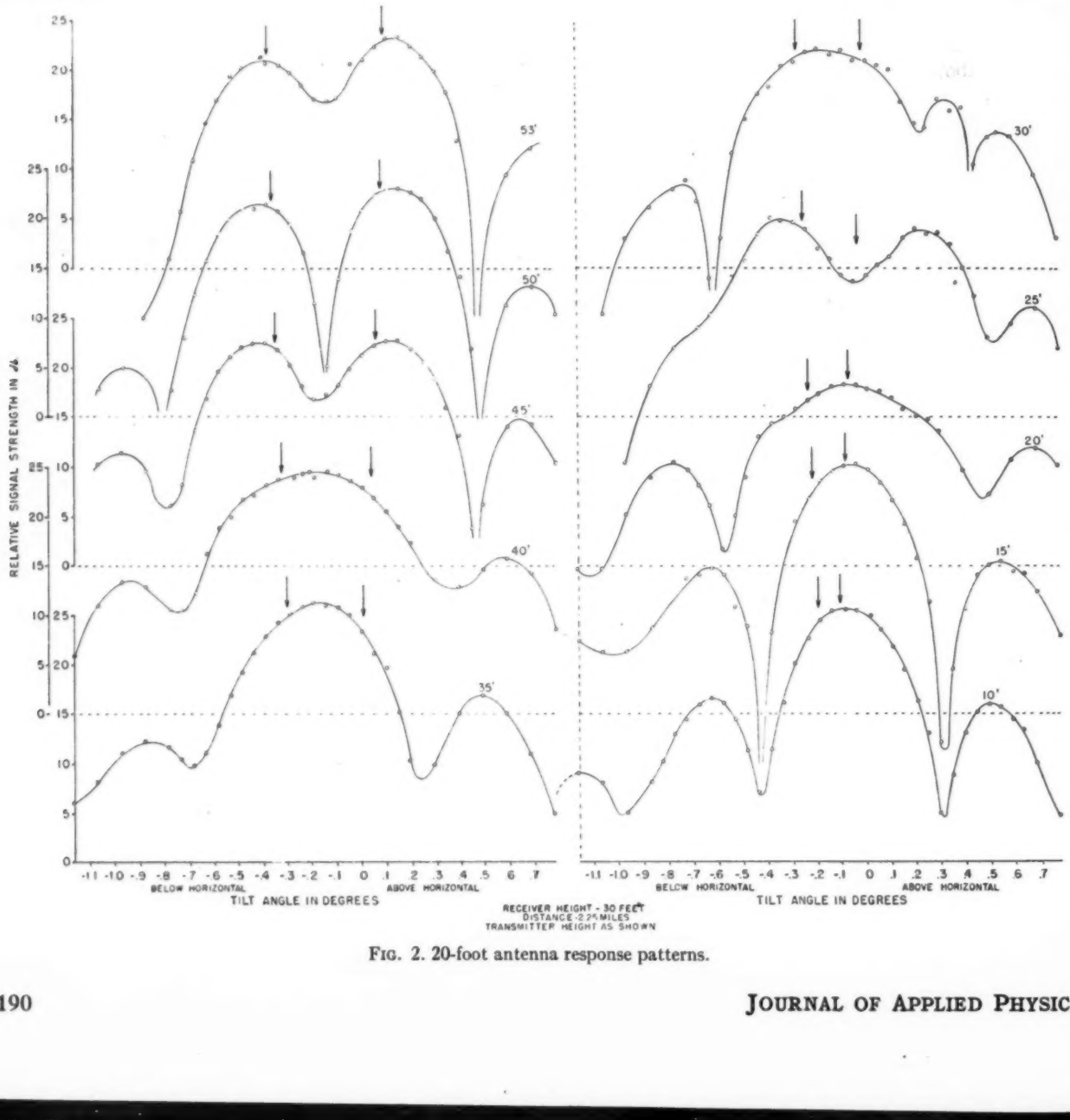

FIG. 2. 20-foot antenna response patterns.

TABLE II. Errors in angles-of-arrival of the direct and reflected rays.* Two-antenna method.

Transmitter height	Separation angle	Direct ray $.00° < \alpha \leq .02°$	$.02° < \alpha \leq .06°$	$.06° < \alpha \leq .10°$	$.10° < \alpha$	Reflected ray $.0° < \rho \leq .02°$	$.02° < \rho \leq .06°$	$.06° < \rho \leq .10°$	$.10° < \rho$
11 feet	0.10°	19	12	1	0	16	15	1	0
21 feet	0.19°	2	2	1 (1)	1 (1)	1	2 (1)	2	1 (1)
33 feet	0.29°	14 (2)	13 (2)	5	0	15 (1)	11 (1)	6 (2)	0
43 feet	0.38°	0	2 (1)	0	2 (2)	1	1 (1)	0	2 (2)
53 feet	0.47°	13	11 (2)	10 (5)	11 (11)	7	22 (5)	5 (3)	10 (10)
		40.7% (50.5%)	33.9% (38.5%)	13.6% (11.0%)	11.9% (0 %)	33.9% (42.9%)	43.2% (47.3%)	11.9% (9.9%)	11.0% (0 %)

* Numbers in parentheses are the number of cases which were rejected. Percentages in parentheses are based on the number of accepted cases.

changes in refraction with meteorological changes were small, linear values for index gradients were assumed for the height-range of the direct and of the reflected rays. The angles of Fig. 1 were corrected by use of the appropriate M-gradient, and the corrected values compared to those measured.

For a qualitative evaluation of the test, the results were grouped into four classes. The first was the group for which the radio measurements differed from the geometrical angles corrected for refraction by 0.02 degree or less. The second group included measurements for which the error was 0.06 degree or less, but greater than 0.02 degree. The third group included those with errors of 0.1 degree or less, but greater than 0.06 degree. The fourth group included those cases for which the error was greater than 0.1 degree. The groups might be classified as (1) extremely accurate, (2) acceptable, (3) inaccurate, and (4) useless.

The results of the 164 cases measured by the three-antenna method are shown in Table I. The division of the measurements between the four groups is shown in percentage in two ways at the foot of Table I. The first percentages are based on all of the measurements made. The second percentage excludes the cases for which the data would indicate that correct results could not be obtained. The number of cases omitted in the second percentage is shown in parentheses in the tabulation.

The excluded cases included several for which the direct and reflected waves are nearly in-phase or 180 degrees out-of-phase at the center antenna. It has been shown previously[4] that the method used becomes very inaccurate for these conditions. These conditions are known from the data to exist when the phase difference between the upper and middle and lower and middle antennas are equal, and the corresponding signal-strength ratios are equal.

All cases for which the ratio of the signal at two antennas exceeded 12 db were aslo excluded since the equipment was known to be inaccurate under these conditions. For a single-component wave arriving at the receiver, the signal-strength ratios will both be one and the phase difference between antenna pairs will be equal to each other. Under these conditions, the solution provided by the three-antenna method will be the true angle-of-arrival of the single-component wave plus a fictitious angle 0.3 degree above the true angle. In the case of the measurements made with the transmitter at 11 feet, the signal strength ratios are nearly one, and the angle between antenna pairs nearly equal to each other. This means that an extremely high degree of accuracy is required in obtaining the data if the angle-of-arrival of each component is to be determined. Apparently, this accuracy was not always met, and the result was that the apparatus measured only a single ray coming from approximately halfway between the transmitter and its reflection. The second angle did not exist but was introduced in the mathematical processes. This accounts for the fact that so many of the direct-ray measurement errors were between 0.02 and 0.06 degree, and the "reflected" ray errors were greater than 0.1 degree when the transmitter was at 11 feet.

The calculation of the angles-of-arrival by the three-antenna method requires subtracting the products of the signal-strength ratios and sines and cosines of the phase differences. When the products to be subtracted are of approximately the same magnitude, accuracy with which the difference is known is very low. For this reason, a very high degree of accuracy in the original data must be obtained. Cases where the difference of the two products was less than 0.1 were placed in the excluded group.

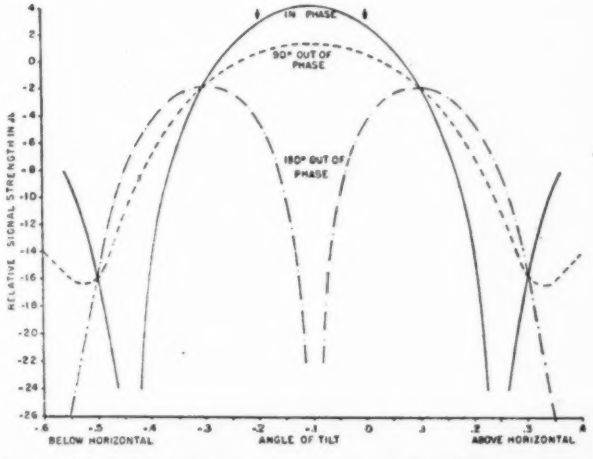

FIG. 3. Theoretical antenna response pattern for 20-foot antenna.

TABLE III. Errors in angles-of-arrival of the direct and reflected rays.* Amplitude method.

Transmitter height	Separation angle	Direct ray $.00° < \alpha \leq .02°$	$.02° < \alpha \leq .06°$	$.06° < \alpha \leq .10°$	$.10° < \alpha$	Reflected ray $.00° < \rho \leq .02°$	$.02° < \rho \leq .06°$	$.06° < \rho \leq .10°$	$.10° < \rho$
11 feet	0.10°	0	0	0	1	0	1	0	0
21 feet	0.19°	0	0	1	2	1	1	0	1
33 feet	0.29°	0	0	0	2	0	0	0	2
43 feet	0.38°	0	0	1	1	0	0	1	1
53 feet	0.47°	1	1	0	0	0	0	2	0
		10%	10%	20%	60%	10%	20%	30%	40%

* Numbers in parentheses are the number of cases which were rejected. Percentages in parentheses are based on the number of accepted cases.

Although the results of this initial field test of the three-antenna method cannot be classed as highly satisfactory, it is felt that with improvements in the accuracy of the original data will come improvements in the accuracy of measuring the angles.

2. Two-Antenna Method

The two-antenna method required as data a height-gain curve and the phase difference between two points. The phase measurements of the three-antenna method provided the necessary phase difference data for the two-antenna method, and a height run after most of the three-antenna tests provided the necessary signal strength data.

One hundred and eighteen sets of measurements were made by the two-antenna method, and the results were classified in the same manner as for the three-antenna method. This classification is shown in Table II. Two sets of percentages are again shown: one for all the data, and one excluding the data for which the signal ratio between the antennas exceeded 12 db.

The height-gain curves agree precisely with those which would be obtained from two waves with a relative magnitude of 0.94. This equivalent reflection coefficient was virtually the same for all days, and was independent of transmitter and receiver heights.

The accuracy of the angle-of-arrival calculations by the two-antenna method is very satisfactory. However, this method has obviously disadvantages which limit its usefulness. A height-gain curve including signal strength maximum and minimum is not always obtainable. In addition, all of the needed data are not available instantaneously.

For fluctuating or changing conditions, this makes the angle determination difficult or impossible. However, the conditions of the present test made possible highly accurate evaluation of the angles-of-arrival by the two-antenna method.

3. Amplitude Method

Scanning patterns for the 20-foot antenna were made for a number of transmitter heights on a day when the atmospheric conditions were known to be nearly standard. These antenna patterns for five-foot intervals of the transmitter are shown in Fig. 2. The arrow in the diagram indicates the angle-of-arrival determined from geometry, corrected for refraction.

The angles-of-arrival as determined from the signal strength maxima are classified in Table III in the same way as for the previous methods.

As shown in a previous report,[7] the scanning patterns for two waves arriving simultaneously is greatly affected by the phase relationship of the two waves. This is illustrated by the theoretical scanning patterns shown in Fig. 3. When the two waves of nearly equal amplitude arrive in phase at the center of the antenna face, a single maximum results, positioned between the true angles-of-arrival. When the two arriving waves are 180 degrees out-of-phase at the center of the two antennas, two distinct maxima appear. However, these maxima indicate an angular separation of the two rays which is too large.

V. SUMMARY

The field test of the three-antenna angle-separating equipment showed that under the most desirable circumstances, the angles-of-arrival of two-wave components may be measured accurately. However, more accuracy is required in the original data if reliable results are to be obtained under all conditions.

The three-antenna system is basically a very significant one because it involves no time delay in taking the data. Because of this instantaneous action of the method, it is felt that further development is justified. Greater accuracy could be obtained by increasing the antenna spacing. This would, however, reduce the angle of ambiguity.

The two-antenna method showed a very high degree of accuracy and gave by far the best results of the three methods considered. However, the requirement of a height-run to determine the signal-strength maximum and minimum limits its usefulness.

The amplitude method appears to be unsatisfactory for separating two rays of nearly equal magnitude with angular separations comparable to the width of the basic antenna pattern. However, the difficulty of ambiguity of the measurement is eliminated in this method.

[7] LaGrone, Hamlin, and Straiton, "The indicated angle-of-arrival by phase front analysis," Report No. 12, Electrical Engineering Research Laboratory, The University of Texas, 1948.

The scanning patterns would not be reliable in a rapidly fluctuating field. The original experiments with the 20-foot antenna did not require the separation of two-wave components so nearly equal or separated by as small an angle as in the test described in this report.

VI. ACKNOWLEDGMENT

The Electrical Engineering Research Laboratory wishes to acknowledge its gratitude to the Bell Telephone Laboratories for the use of the parabolic antenna which provided the comparison test.

Immobility Phenomena and Reverse Driving Phenomena of the Electric Arc

SAKAE YAMAMURA

The Electrical Department, Faculty of Engineering, University of Tokyo, Tokyo, Japan

(Received June 22, 1949)

When an electric arc is driven by a magnetic field, under special conditions, curious phenomena are observed. An electric arc cannot be driven by the magnetic field when it is established at very small gaps between the electrodes. Neither can it be driven by the magnetic field for the instant just after the contacts have been broken. These phenomena are called immobility phenomena. In the present work, the immobility time and the driving velocity of the electric arc were measured under various air pressures. As the air pressure decreases, the immobility time increases and the driving velocity decreases. Under a certain low air pressure, the arc cannot be driven by the magnetic field. When the air pressure is lower than this critical value, the arc can be driven again by the magnetic field; but in this case, the arc moves in a direction that is opposite to the direction of the electromagnetic force. This phenomenon is called the reverse driving phenomenon. The causes for these phenomena are considered.

I. INTRODUCTION

WHEN an electric arc is established in the presence of the magnetic field, curious phenomena are observed. Under normal conditions, the arc moves in the same direction as the electromagnetic force. Under some special conditions, such as at a very small gap,[1] just after the contacts have been broken,[1,2] and at low air pressures,[3,4] the arc cannot be driven by the magnetic field. Scars on the electrodes made the electric arc remain stationary. Under still other conditions, the arc moves in a direction that is opposite to the direction of the electromagnetic force.

These phenomena were observed independently by several persons including the author.[1,3] But interrelations between these phenomena have been missed. Some of the experimental results and considerations about these phenomena will be presented in the following reports.

II. EXPERIMENTAL APPARATUS

A picture of the experimental apparatus that was used is shown in Fig. 1. The electric arc is established by separating two parallel copper electrodes. The electrodes are placed horizontally and are sealed in a glass vessel which is connected to a rotary vacuum pump. The driving magnetic field is applied perpendicularly to the arc as is shown by the arrow H in Fig. 1. The electrodes are pulled apart and shorted together by an electromagnet (there is no relation between this magnet and the driving magnetic field). The magnet which pulls apart and shorts together the electrodes is controlled by a rotary switch that is connected to the shaft of a rotating mirror. The image of the arc is focused through a lens and is reflected on the screen by a rotating mirror. Since the mirror is rotating, the image of the arc will repeatedly sweep from the bottom to the top of the screen. The rotating mirror revolves at a constant velocity. In this way, the loci of the images of several successive electric arcs will appear in the same position on the screen. If the arc is stationary, the locus of its image on the screen is a vertical line (as from a to b). If the arc is driven by the magnetic field, the locus of its image is an oblique line (as from b to c). The horizontal velocity of the arc can be determined by measuring the angle θ of the locus on the screen.

III. OBSERVATIONS

1. The Magnetic Driving of a Contact-Arc

An electric arc which is generated by breaking contacts is named a contact-arc. The movements of a contact-arc which was driven by the magnetic field were observed with the apparatus described in Section II. The locus of the image of the contact-arc was the same as the locus shown on the screen in Fig. 1. This locus indicates that the contact-arc is stationary for an instant just after the contacts have been broken; then it begins suddenly to be driven by the magnetic field. After being driven for a moment, the arc goes out. This phenomenon was named the immobility phenomenon.[1,2] The author proposes to name it the momentary immobility phenomenon to distinguish it from the permanent immobility phenomenon. The permanent immobility phenomenon will be discussed later.

[1] S. Yamamura, J. Inst. Elec. Eng. Japan **61**, 660 (1941).
[2] S. Fukuda, J. Inst. Elec. Eng. Japan **62**, 161 (1942).
[3] S. Yamamura, Report of Annual Meeting of Inst. Elec. Eng. Japan, 16 (1948).
[4] G. J. Himler and G. L. Cohn, Elec. Eng. **67**, 1148 (1948).

The immobility time and the driving velocity of the contact-arc were measured under various air pressures. Figure 2 shows one example of these measurements.

As the air pressure decreases, the driving velocity decreases and the immobility time increases. At a certain pressure (in Fig. 2, 350 mm Hg), the electric arc cannot be driven by the magnetic field, and it remains permanently stationary. This phenomenon is called the permanent immobility phenomenon.

When the pressure is reduced below this critical value (350 mm Hg), the arc can be driven again. But

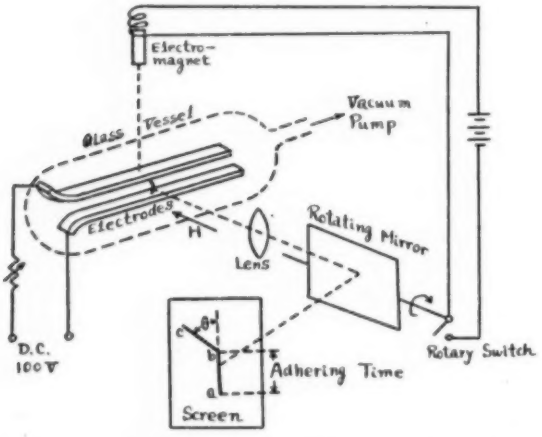

FIG. 1. Apparatus for experimenting with the magnetic driving of the electric arc.

in this case, the arc moves in a direction that is opposite to the direction of the electromagnetic force. This phenomenon is called the reverse driving phenomenon.

As the air pressure decreases further, the reverse driving velocity increases greatly and the immobility time decreases to almost a zero value, as one can determine from Fig. 2.

The same sequence of phenomena occurs when the arc current and the driving magnetic field are of different values; but, in these cases, the critical air pressure at which permanent immobility occurs is different. These relations are shown in Fig. 3. The critical air pressure (the pressure at which permanent immobility occurs) increases as the magnetic field intensity increases. This is contrary to the statements of Himler and Cohn,[4] who reported that the critical air pressure decreased as the magnetic field increased.

The same experiments were performed with several different electrode materials to decide whether they would show permanent immobility and reverse driving. One example using aluminum electrodes is shown in Fig. 4. The aluminum electrodes have a wider pressure range where permanent immobility occurs.

All of the experimental results are summarized in Table I. The materials are arranged as grouped by Druyvesteyn and Penning.[5] These authors arranged the various metals into nine groups according to the mobility of the cathode spot of the metal.

Permanent immobility was observed with all of the materials that were examined. When the air pressure was reduced to 50 mm Hg, reverse driving was observed in all of the groups except groups I and II.

2. Some Other Immobility Phenomena of the Electric Arc

When an electric arc is established at a very small gap between the two parallel electrodes, it cannot be driven by the magnetic field. In this case, even under a normal atmospheric pressure, permanent immobility occurs.

Since a molten metal bridge exists between the electrodes which are separated very slowly, the arc does not appear until the electrodes have been separated to the displacement (D_1) which is shown by the broken line in Fig. 5. The continuous line of Fig. 5 indicates the displacement (D_2) of the movable electrode at which the arc begins to move quickly. Between displacements (D_1) and (D_2) the arc cannot be driven by the magnetic field.

Scars on the electrodes, especially on the cathode, also make the electric arc remain stationary.

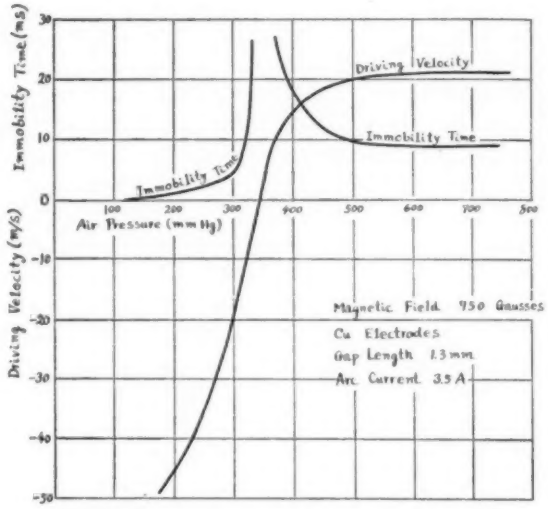

FIG. 2. The relations between the air pressure, the immobility time and the driving velocity of the contact-arc.

IV. CONSIDERATIONS

C. G. Smith[6] reported about the reverse driving of the cathode spot of a mercury arc. He explained this phenomenon by using the negative Righi-Leduc effect, but his explanation does not seem plausible. Not only do the metals with a negative coefficient of the Righi-Leduc effect (for instance Mg, Cu and Ag), but also the metals with a positive coefficient (for instance Fe) show reverse driving. It is impossible that the coeffi-

[5] M. J. Druyvesteyn and F. M. Penning, Rev. Mod. Phys. **12**, 87 (1940).

[6] C. G. Smith, Phys. Rev. **62**, 48 (1942).

TABLE I. Classification of the electrode materials according to the characteristics of the arc on them.

	I	II	III	IV	V	VI	VII	VIII	IX
Electrode material	pure C	pure solid Zn pure solid Pb	normal Cu normal Fe normal Ag	Al Sn	impure Fe	very pure Cu very pure Fe very pure Ag	liquid Hg*	normal Cu	(Al) (Sn)
Impurities present	no	no	yes	yes	yes	no	no		
Permanent immobility	yes	yes	yes	yes	yes	yes	yes	yes	yes
Reverse driving	no	no	yes	yes	yes	yes	yes	yes	yes

* According to C. G. Smith (see reference 6).

cient of the Righi-Leduc effect changes its sign from positive to negative with any change of air pressure.

Our explanation involves the electron diffusion theory. Himler and Cohn[4] and the author[3] applied almost the same explanation. An enlarged picture of a cathode spot is shown in Fig. 6. When the magnetic field is applied perpendicularly from the front to the back of the paper, the loci of electrons leaving the cathode are bent to the right side by the magnetic field as is shown in Fig. 6. Thus there ensues the rightward movement of the arc, which is the normal magnetic driving of the arc.

When the atmospheric pressure is much lower, that is, when the mean free path of the electrons is longer, the path of the electrons appears to be influenced to a great degree by the magnetic field. Many electrons with energy corresponding to the cathode fall, diffuse outwards in the positive space charge cloud. The loci of these electrons are bent by the transversal magnetic field as is shown in Fig. 6. The electrons diffusing to the left side are accelerated by the electric field, and thus these electrons efficiently ionize the gas molecules by collision. The electrons diffusing to the right side are decelerated by the electric field because these electrons are moving in an opposite direction to the electric field. Thus the electrons moving to the right side cannot efficiently ionize the gas molecules by collision. In this way, the positive space charge cloud spreads to the left side.

If this leftward spreading of the positive space charge cloud overcomes the rightward spreading (the normal magnetic driving of the arc), then reverse driving ensues, If these two spreadings balance each other, then permanent immobility ensues. These phenomena can occur only when the diffusing electrons have a comparatively long mean free path, and when their loci are bent notably by the magnetic field. Let us check this.

From Fig. 3 we can see that the critical air pressure at which permanent immobility occurs is 100–200 mm Hg at 360 gausses. The electron mean free path (λ_0) in the atmosphere is 3.6×10^{-5} cm at 0°C. The temperature of the positive column of the arc is 5000–10,000°K. A great many electrons in the positive space charge cloud, which have been accelerated by the cathode fall, have a very high temperature. Since this is true, the temperature of the gas molecules in the positive space charge cloud would be higher than the temperature of the positive column. Therefore the temperature of the gas molecules could be assumed to be 10,000°K. Then the electron mean free path (λ_e) at 100 mm Hg is

$$\lambda_e=\lambda_0\times\frac{760}{p}\times\frac{T}{273}=3.6\times10^{-5}\times\frac{760}{100}\times\frac{10{,}000}{273}=1.0\times10^{-2}\text{ cm}. \quad (1)$$

The radius R of a circle, which an electron with the energy of V electron-volts follows under the influence of the perpendicular magnetic field of H gausses, is

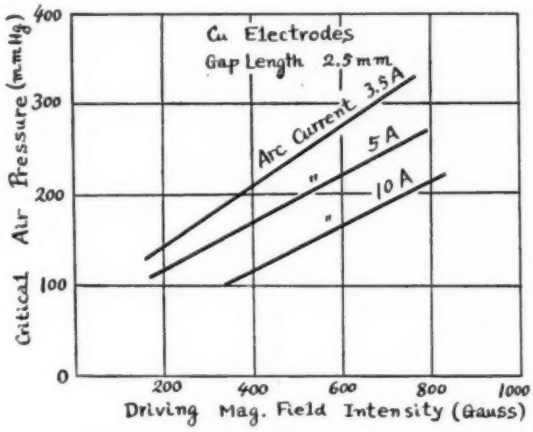

FIG. 3. Relations between driving magnetic field intensity and critical air pressure at which the permanent immobility occurs.

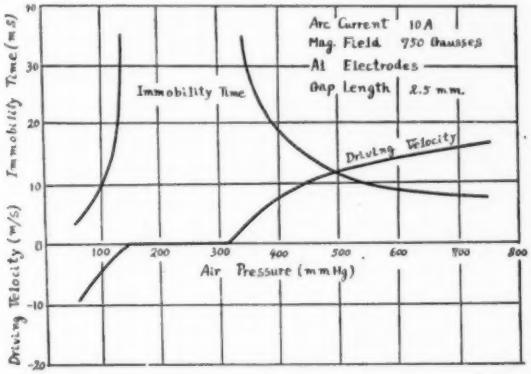

FIG. 4. The same relations as in Fig. 2. But the electrodes are made of aluminum.

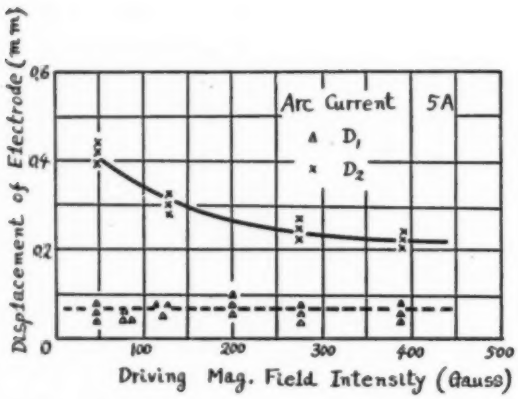

FIG. 5. Relations between driving magnetic field intensity, and electrode displacement D_1 at which an electric arc appears, and electrode displacement D_2 at which an electric arc begins to be driven. The electrodes are clean.

given by the formula

$$R=\frac{3.4(V)^{\frac{1}{2}}}{H}\text{ cm.} \tag{2}$$

Many of the electrons diffusing outward in the positive space charge cloud have been accelerated by the cathode fall before they make an elastic collision with a gas molecule. The energy of the diffusing electrons (the electrons that have left the cathode fall) may be assumed to be 8 electron-volts. This corresponds to the cathode fall of the copper electrode. When H is 360 gausses, then

$$R=\frac{3.4(8)^{\frac{1}{2}}}{360}=2.7\times10^{-2}\text{ cm.} \tag{3}$$

The electron mean free path and the radius are nearly equal. Thus it is probable that the loci of the many diffusing electrons are bent notably by the magnetic field as shown in Fig. 6. Although it is accompanied by some uncertainties due to complications in the cathode spot, this calculation would confirm the electron diffusion theory of reverse driving and of the immobility phenomena under a low air pressure.

As the driving magnetic field increases, the radius R of the circular loci of the electrons given by Eq. (2) becomes smaller. Thus permanent immobility will occur at a higher air pressure with an increase in the driving magnetic field.

Momentary immobility of the electric arc might be due to the fact that the temperature of the cathode spot might be very high when the arc is first established. Thus for an instant, the electron mean free path is relatively long and much curved by the magnetic field. Permanent immobility at very small gaps may be attributed to the same reason. That is, the gas temperature is very high.

Easily vaporizable metals (groups III, IV in Table I) have a relatively long time of momentary immobility. Very easily vaporizable metals (group II in Table I) have a very long time of momentary immobility and show no reverse driving. Materials with a thermoemissive cathode spot (group I in Table I) show no reverse driving.

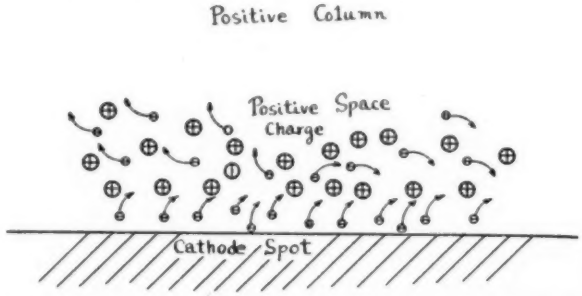

FIG. 6. Enlarged picture of the cathode spot. ⊕ positive ions. ⊖ electrons. The magnetic field is directed perpendicularly from the front to the back of the paper.

V. CONCLUSION

From the characteristics of the immobility phenomena and reverse driving phenomena which were reported, it is evident that they must be taken into consideration in the magne-blast circuit breaker. Momentary immobility not only would lengthen the interrupting time of the breaker, but it would also damage the most important parts of the contacts. Permanent immobility would make the breaker lose its interrupting ability at an altitude of several thousand meters. An air blast could protect the breaker from these phenomena effectively.

There remain many matters to prove experimentally. Further experiments and details of practical considerations mentioned will be reported later.

The author expresses his sincere thanks to Professors H. Yamashita, S. Fukuda, and S. Hoh for their guidance.

The author thanks sincerely Mr. J. P. LeGeros for his kind help with the manuscript. The author also thanks Mr. M. Yasui very much for his effective assistance.

Measurements of Mechanical Properties of Polyisobutylene at Audiofrequencies by a Twin Transducer*

ROBERT S. MARVIN,† EDWIN R. FITZGERALD,‡ AND JOHN D. FERRY
Departments of Chemistry and Physics, University of Wisconsin, Madison, Wisconsin
(Received July 25, 1949)

Apparatus has been developed for measuring the dynamic viscosity and rigidity of soft rubber-like solids in small oscillating deformations. A plate rigidly attached between two identical coils in two permanent magnets shears a pair of disk-shaped samples when a driving current is passed through one coil. The open circuit voltage from the other coil is compared in amplitude and phase with the driving current by a method in which all measurements are in the form of settings of a potential divider. The apparatus has two advantages over the more familiar resonance devices: (a) the amplitudes of motion, which need not be measured directly, are extremely small, thus minimizing any non-linear effects, or temperature change due to heat dissipation; (b) a continuous range of frequencies, spaced as closely as desired, is available without adjusting masses. The dynamic rigidity and viscosity of two samples of polyisobutylene, of molecular weights 1.2 and 0.47 million, have been measured at 15, 25, and 35°C at frequencies from 20 to 600 cycles/sec., with a reproducibility of five percent or better. The rigidity increases and the viscosity decreases with increasing frequency. Both decrease with increasing temperature, and are almost independent of molecular weight. The temperature dependence is quantitatively explained by the assumption that stresses relax by flow processes whose apparent activation energies are all identical with that characterizing the steady flow viscosity.

INTRODUCTION

THE interpretation of equilibrium rubber-like elasticity in terms of a three-dimensional network is well established.[1,2] In agreement with this theory, uncrosslinked linear polymers such as raw rubber and polyisobutylene have no equilibrium elasticity; when a sample is held at constant strain, the stress decays to zero.[3] Nevertheless, these unvulcanized rubbers show elasticity under dynamic or non-equilibrium conditions;[4] in a polyisobutylene of molecular weight about one million, the stress-strain ratio 12 sec. after application of stress[3] corresponds to an elasticity modulus of about 5×10^6 dyne/cm^2, and dynamic measurements at 100 cycles/sec. correspond to a dynamic Young's modulus of about 12×10^6 (see below), as compared with equilibrium values of the order of 10×10^6 for typical vulcanized rubbers. The origin of this non-equilibrium elasticity is not well understood. Knowledge of the phenomenon can be advanced by quantitative measurements of mechanical properties in small oscillating deformations.

Although many experimental methods for measuring dynamic properties of rubber-like solids have been described recently,[5–7] most are not very well adapted for study, over a range of audiofrequencies, of uncrosslinked materials which cannot support static stress even for relatively brief periods. For measurements on such soft polymers, as well as on swollen gels and highly concentrated solutions, we have constructed a twin electromagnetic transducer. This instrument has the advantages that the shear strains employed are very small (of the order of 10^{-5} to 10^{-3}), measurements can be made over a continuously variable frequency range (not being restricted to operation at mechanical resonance), and all the data are obtained in a particularly simple and precise form—as settings of a potential divider.

DESCRIPTION OF APPARATUS

Twin voice coils, wound on Lucite forms, are joined by short brass tubes to a brass plate, about one inch square (Fig. 1), and are suspended by eight piano wires, diameter 0.007 in., attached to small plates soldered to the tubes. The wires are adjusted under moderate tension so the coils are accurately centered in the air gaps of twin loudspeaker magnets (Jensen[8] Model PMB S3575). Identical disk-shaped samples can be strained in shear between the central brass plate and the ends of two fixed brass cylinders. The latter are held by a split steel block, also bolted to the base plate, and can be clamped by tightening two screws through the block, one of which is shown. Before clamping, the cylinders are advanced by calibrated lead screws, and the distance between cylinder and plate, representing the thickness of the sample, can be read on each side on circular scales.

A photograph of the apparatus with one magnet removed is shown in Fig. 2. The base plate rests in a box of sawdust on a concrete block mounted on cork. For temperature control, the apparatus is covered by a Celotex box connected by a rubber bellows to an air thermostat. The temperature of the central steel block, measured by a thermometer in an oil-filled well, is controlled to ±0.3°C or better. Tests with a thermocouple have shown that under the conditions of our experiments (which involve extremely

* Presented in part at a meeting of the American Physical Society, Madison, Wisconsin (June 21, 1948) and at a meeting of the Rubber Division of the American Chemical Society, Los Angeles (July 22, 1948).

† Carbide and Carbon Chemicals Corporation Fellow in Physical Chemistry (1947–49). Present address: Rubber Section, National Bureau of Standards, Washington 25, D. C.

‡ Department of Physics.

[1] Guth, James, and Mark, Adv. Colloid Sci. **2**, 253 (1946).

[2] P. J. Flory, Ind. Eng. Chem. **38**, 417 (1946).

[3] Andrews, Hofman-Bang, and Tobolsky, J. Polymer Sci. **3**, 669 (1948).

[4] Cf. K. H. Meyer and A. J. A. Van der Wyk, J. Polymer Sci. **1**, 49 (1946).

[5] A. W. Nolle, J. App. Phys. **19**, 753 (1948).

[6] J. H. Dillon and S. D. Gehman, Ind. Rubber World **115**, 16, 217 (1946).

[7] W. Kuhn and O. Künzle, Helv. Chim. Acta **30**, 839 (1947).

[8] Jensen Manufacturing Company, Chicago, Illinois.

small amplitudes of oscillation and hence very little heat dissipation) the temperature of the sample and the block are identical within 0.05°.

When an alternating e.m.f. is applied to one coil, the samples are subjected to sinusoidally varying shear strain, and an e.m.f. is generated in the second coil. To calculate the dynamic mechanical properties of the samples from transducer relationships, it is necessary to compare the applied and generated voltages in amplitude and phase. For this purpose, the electrical circuit shown in Fig. 3 is employed. The oscillator is a Hewlett-Packard[9] Model 200 CR. Four non-inductive resistors in series form a potential divider represented by R_1 and R_2, which is followed by a three-dial decade voltage divider R_2, General Radio[10] Type 654A, with a total impedance of 10,000 ohms. If we consider the input impedance at *VTVM* as infinite (it is actually 0.5 megohm, and this approximation involves a maximum error of 0.5 percent) the potential E_1 appearing at the output of the decade box R_2 will be αRI_1, where α is the fractional setting of the decade box, $1/R=1/R_1+1/R_2$, and I_1 is the current through the driving coil. The values of the resistors composing R_1 and R_2 were chosen so that R would be exactly 1, 10, or 100 ohms, plus a fourth resistance of about 1000 ohms to give the proper load across the oscillator. The potential E_1 is opposed to the e.m.f. of the driven coil, $\mathbf{E}_2$, and the resultant $\mathbf{E}_r$ is read on a Ballantine[11] Model 302 vacuum tube voltmeter, preceded by a Model 220 amplifier.[11] The high input impedance of the amplifier insures that no appreciable current is drawn from the driven coil. By means of a switch, E_1 may also be read separately. All connections are carefully shielded.

THEORY AND CALCULATIONS

From the principles of electromagnetic transducers[12] it can be readily shown that

$$\mathbf{E}_2=(B_1l_1B_2l_2\times10^{-9}/\mathbf{Z_M})I_1, \tag{1}$$

where B_1 and B_2 are the magnetic field strengths of the speakers, l_1 and l_2 the lengths of wires in the coils, and $\mathbf{Z_M}$ the mechanical impedance defined as the ratio of the driving force on coil 1 to the velocity of coil 2. The real part of $\mathbf{E}_2$ is taken in phase with I_1. The voltmeter provides only the absolute value, $|\mathbf{E}_2|$. However, its components can be separated by opposing it to E_1 (which is in phase with I_1 and therefore written as real) as described and using the following procedure.

Denoting $B_1l_1B_2l_2\times10^{-9}$ by K^2, and using the admittance $\mathbf{Y_M}=1/\mathbf{Z_M}=Y_M'-iY_M''$, we have $\mathbf{E}_2=K^2(Y_M'-iY_M'')I_1$, and the resultant voltage $\mathbf{E}_r$ is

$$\mathbf{E}_r=E_1-\mathbf{E}_2=[\alpha R-K^2Y_M'+iK^2Y_M'']I_1. \tag{2}$$

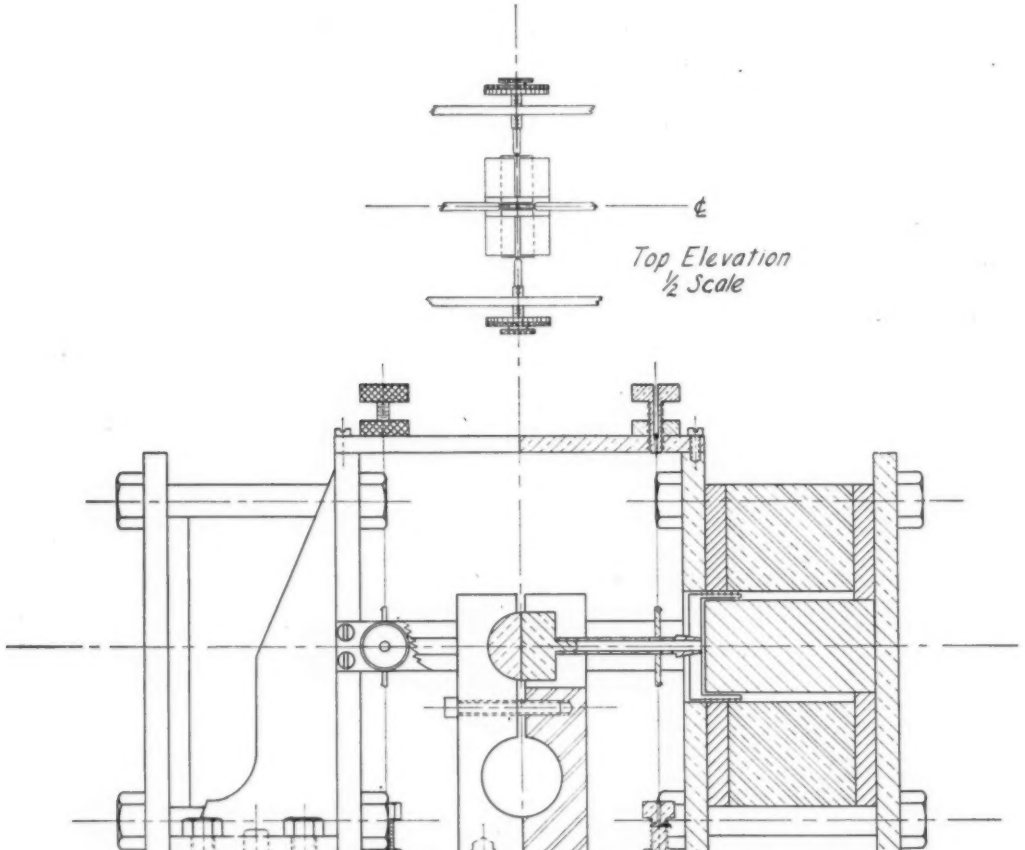

FIG. 1. Quarter-section drawing of twin transducer. Base plate is 12 in. long.

[9] Hewlett-Packard Company, Palo Alto, California.
[10] General Radio Company, Cambridge, Massachusetts.
[11] Ballantine Laboratories, Inc., Boonton, New Jersey.
[12] W. P. Mason, *Electromechanical Transducers and Wave Filters* (D. Van Nostrand and Company, Inc., New York, 1942).

When the potential divider setting α is varied at constant driving current and frequency, $|\mathbf{E}_r|$ goes through a minimum when the first two terms in the coefficient of I_1 cancel. At this point, the divider setting determines K^2Y_M' and the voltmeter reading determines K^2Y_M''; the coil e.m.f. $|\mathbf{E}_2|$ provides a check calculation:

$$K^2Y_M' = (\alpha R)_{\min}$$
$$K^2Y_M'' = |\mathbf{E}_r|_{\min}/I_1$$
$$K^2(Y_M'^2+Y_M''^2)^{\frac{1}{2}} = K^2|\mathbf{Y_M}| = |\mathbf{E}_2|/I_1. \qquad (3)$$

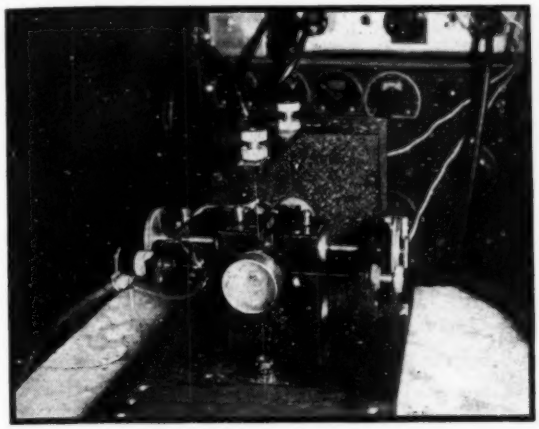

FIG. 2. Photograph of twin transducer with one magnet removed.

Since only ratios of voltage to current are required, the voltmeter scale readings for $|\mathbf{E}_r|$ and $|\mathbf{E}_2|$ are in practice matched by adjusting the potential divider with $\mathbf{E}_2$ switched out, at constant I_1, and the ratios $|\mathbf{E}_r|/I_1$ and $|\mathbf{E}_2|/I_1$ are thus obtained directly without relying upon calibration of the voltmeter scale. The check given by the last of Eqs. (3) is within one percent unless fluctuation in I_1 or frequency has occurred during the measurements.

The minimum is rather shallow; in some cases its location is facilitated by placing a resistance in series with the d.c. output of the vacuum tube voltmeter, and measuring the voltage drop across this resistance with an ordinary potentiometer and sensitive galvanometer.

The mechanical properties of the material subjected to sinusoidally varying shear are described by the dynamic rigidity, G' (real part of the complex dynamic shear modulus), and the dynamic viscosity, η' (real part of the complex dynamic viscosity), which have been defined elsewhere.[13,14] If the mass of the samples can be neglected, or their thickness is small compared to the wave-length of shear waves in the material, these quantities are related to the mechanical admittance as follows:

$$Y_M'/|\mathbf{Y_M}|^2 = R_M = R_M^0 + 2(A/h)\eta' \qquad (4)$$
$$Y_M''/|\mathbf{Y_M}|^2 = X_M = X_M^0 - 2(A/h)G'/\omega \qquad (5)$$

where R_M and R_M^0 are the mechanical resistances of the moving system with and without samples, X_M and X_M^0 are the corresponding mechanical reactances, A and h are the area and thickness of the two identical disk-shaped samples, and ω is the circular frequency. If the propagation of an incipient shear wave in the sample is taken into account, the contribution of the samples to the mechanical impedance is given by a hyperbolic cotangent function[15] which can be expanded to give a better approximation replacing Eq. (5):

$$Y_M''/|\mathbf{Y_M}|^2 = X_M = X_M^0 - 2(A/h)G'/\omega + 2\omega m/3 \qquad (6)$$

[13] Ferry, Sawyer, and Ashworth, J. Polymer Sci. **2**, 593 (1947); cf. H. Leaderman, J. Colloid Sci. **4**, 193 (1949).

[14] Smith, Ferry, and Schremp, J. App. Phys. **20**, 144 (1949).

[15] R. S. Marvin, Ph.D. thesis, University of Wisconsin (1949).

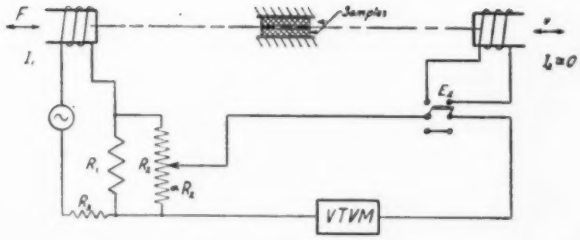

FIG. 3. Electrical circuit used with transducer.

where m is the mass of each sample. In practice, Eqs. (4) and (6) are employed, although the contribution of the term in m is slight (cf. Table I).

CALIBRATION

At each frequency, three apparatus constants are required: K^2, R_M^0, and X_M^0. The ratios R_M^0/K^2 and X_M^0/K^2 are readily determined from measurements without samples, using Eqs. (3)–(5). These values are plotted against the frequency in Fig. 4. Some isolated determinations showed that they were the same at 25° and 35°C.

The above discussion has tacitly assumed that the contributions of samples and apparatus to mechanical resistance and reactance are simply additive, with no interaction. If the coils, tubes, and plate moved as a rigid unit, this would be the case, and, moreover, X_M^0 would be equal to $\omega M - s_1/\omega$, where M is the mass of the coil-tube-plate assembly (68.2 g), and s_1 the elastance of the supporting wires. At frequencies well above the resonance $\omega_0 = (s_1/M)^{\frac{1}{2}}$, which corresponds to $\nu \cong 24$ cycles/sec., X_M^0 should be simply ωM. Figure 4 shows that this is a good approximation up to about 200 cycles/sec.; however, at higher frequencies X_M^0/K^2 falls off from the $\omega M/K^2$ line, goes through a maximum, and falls to zero at about 3000 cycles/sec. (not shown). This behavior is attributed to a mode of vibration in which the coils move in and out symmetrically about a node at the plate, owing to compliance of the tubes, with a resonance frequency of about 3000 cycles/sec.[16] Because of this complication, no measurements of G' have been reported above 600 cycles/sec. with the present apparatus. Up to this frequency, the additivity expressed in Eqs. (4)–(6) has been assumed; its validity is shown by the agreement in G' and η' calculated from measurements on samples of different sizes (see below).

The constant K^2 is obtained by plotting $\omega Y_M''/|\mathbf{Y_M}|K^2(=\omega X_M^0/K^2)$, determined from measurements without samples, against ω^2. If the coil-tube-plate assembly moves as a rigid unit, this should give a straight line with a slope of M/K^2 and an intercept of $-s_1/K^2$. From 20 to 100 cycles/sec. an excellent line is obtained from whose slope K^2 is found to be 8.5×10^4 ohm-dyne-sec.-cm^{-1}. This value is considered to be independent of frequency.[17, 17a]

[16] There are other indications that, even at fairly low frequencies, the mechanical impedance does not correspond to simple motion of a rigid rod, for which X_M would be $\omega M - S_M/\omega$, where $S_M = s_1 + 2(A/h)G'$. (1) There is a sharp discontinuity in the slope of the X_M^0/K^2 calibration curve at about 680 cycles (Fig. 5). (2) Values of G' obtained by applying the simple equation above to the value of the resonance frequency observed with a sample are about ten percent higher than those obtained from Eqs. (3) and (6). (3) If very thin samples are employed, two or even three resonances are observed below 1000 cycles, indicating strong interaction with the high frequency resonances. No such samples were used in obtaining the data reported here.

[17] A more elaborate analysis of calibration measurements at high frequencies, in which the compliance of the tubes is taken into account, shows that K^2 is independent of frequency at least up to 1500 cycles.

[17a] A recent method for the absolute calibration of electromechanical transducers [H. M. Trent, J. App. Mech. **70**, 49 (1948); S. P. Thompson, J. Acous. Soc. Am. **20**, 637 (1948)] has

TABLE I. Measurements and calculations of dynamic mechanical properties of polyisobutylene, $\bar{M}_\eta = 1.2\times10^6$, at 15°C; $m = 0.163$ g, $2A/h = 20.3$ cm.

Frequency, cycles/sec.	R_M/K^2 ohm^{-1}	$R_M{}^0/K^2$ ohm^{-1}	$\eta'\times10^{-3}$ poises	X_M/K^2 ohm^{-1}	$X_M{}^0/K^2$ ohm^{-1}	$2\omega m/3K^2$ ohm^{-1}	$G'\times10^{-4}$ dyne/cm^2
30	3.533	0.04	14.61	−4.932	0.055	0.000	3.94
45	3.069	0.04	12.68	−3.581	0.163	0.000	4.43
60	2.793	0.04	11.53	−2.840	0.258	0.000	4.89
90	2.424	0.05	9.94	−1.974	0.428	0.001	5.69
150	2.008	0.07	8.11	−1.022	0.741	0.001	6.96
200	1.812	0.08	7.25	−0.552	0.980	0.002	8.06
240	1.686	0.08	6.72	−0.211	1.160	0.002	8.66
300	1.514	0.08	6.00	+0.242	1.421	0.002	9.31
400	1.343	0.08	5.29	0.871	1.885	0.003	10.67
500	1.247	0.08	4.89	1.563	2.270	0.004	9.31
600	1.111	0.08	4.32	1.971	2.660	0.005	10.9
700	0.820	0.07	3.14	2.383	3.160	0.006	—
800	0.928	0.07	3.59	2.809	3.420	0.006	—
900	0.830	0.06	3.23	3.222	3.660	0.007	—
1200	0.660	−0.01	2.80	4.007	4.155	0.010	—
1500	0.349	−0.30	2.72	4.164	4.170	0.012	—

RESULTS ON POLYISOBUTYLENE

Two unfractionated samples of polyisobutylene were studied at 15°, 25°, and 35°C. The first (viscosity-average molecular weight 1.2 million[18]) was furnished through the kindness of Dr. John Rehner, Jr., Esso Laboratories. The second (viscosity-average molecular weight 0.47 million[18]), originally from Dr. Rehner's laboratory, was kindly given us by Dr. Rodney D. Andrews, Princeton University. Portions were compressed to sheets of uniform thickness (the relaxation of stress hastened by warming to 60°), and disks were cut by dies. In each case the disk was weighed and its volume was calculated, taking the density as 0.91. The volume divided by the square of the thickness after compression (determined by the cylinder lead screws in the apparatus, and often checked in place with a thickness gauge) gives the ratio A/h.

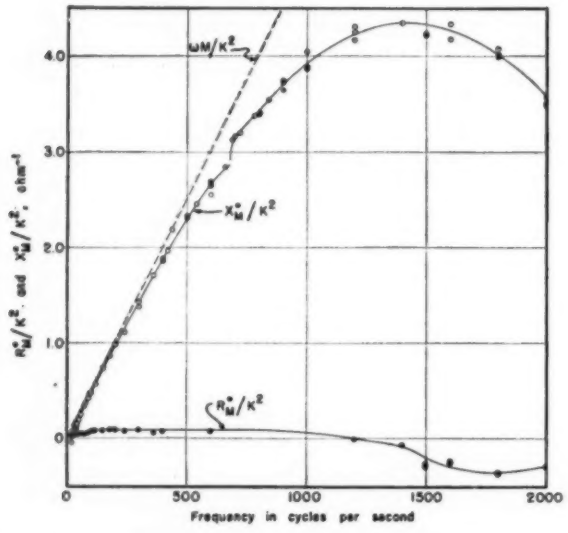

FIG. 4. $R_M{}^0/K^2$ and $X_M{}^0/K^2$ plotted against frequency, at 25°C. The dashed line represents $\omega M/K^2$.

To insure contact between the samples and the plate and cylinders, the disks were usually compressed about ten percent after introduction, and allowed to stand several hours to permit the compressive stress to relax. Tests showed that the dynamic properties did not change during this relaxation, however. The polyisobutylene was sufficiently tacky to make excellent mechanical contact following this procedure; in fact, it was not easy to remove the disks after completion of measurements.

Table I gives an example of the magnitudes of R_M and $R_M{}^0$ and of η' calculated from their difference, as well as the corresponding quantities X_M, $X_M{}^0$, and G'. The magnitude of $R_M{}^0$ is never large enough so that its subtraction seriously affects the accuracy of η'. At low frequencies, $X_M{}^0$ is not a serious correction, but with increasing frequency its contribution to X_M of course becomes greater until their difference becomes too small to determine G' with adequate precision.

Values of G' and η' for the high molecular weight sample at three temperatures are plotted against the logarithm of the frequency in Figs. 5 and 6, and the same quantities for the low molecular weight sample in Figs. 7 and 8. In each case, two different sets of disks have been employed, with A/h ratios differing by a factor of two to four. The agreement of the values of G' and η' obtained, shown by the tagged points, justifies the use of Eqs. (4) and (6) within the frequency range employed.

Individual readings on the same sample are generally reproducible within about one percent, giving an error of about two percent in the values of R_M and X_M. Values for different samples show slightly more deviation, probably reflecting an additional error in the

the advantage that the compliances of coils and connector are included with K^2 in a single calibration which is of course a function of frequency. No assumption of additivity of impedances is necessary. The method requires addition of masses for calibration at the center of the connector; this is not possible with our present apparatus, whose construction was completed before the appearance of Trent's work.

[18] We are indebted to Dr. J. N. Ashworth and Mr. M. F. Johnson for the molecular weight values.

measurement of h, the sample thickness. Very few values for the pairs of samples shown in Figs. 5–8 are more than five percent apart, and this seems to be a safe figure to take for the reproducibility of the values reported here.

These data were obtained with a constant driving current of 10 ma; the peak amplitude and velocity of course varied depending on the total mechanical impedance and the frequency. The shear and rate of shear were always so small, however, that linear viscoelastic behavior[13,14] is to be expected, and the quantities measured should be independent of amplitude. In several cases this was confirmed by varying I_1 over a factor of four, with no resultant change in R_M or X_M.

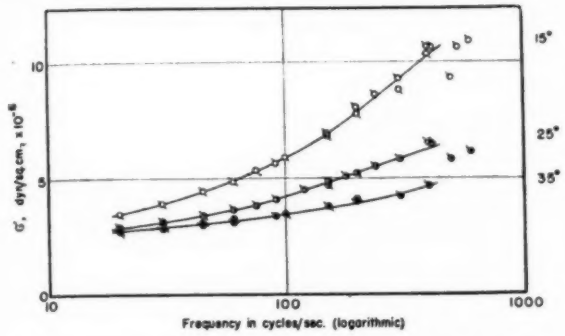

FIG. 5. Real part of dynamic rigidity of polyisobutylene, $\bar{M}_\eta = 1.2 \times 10^6$, plotted against the logarithm of the frequency. Upward tags, $2A/h = 20.3$; downward tags, $2A/h = 5.20$.

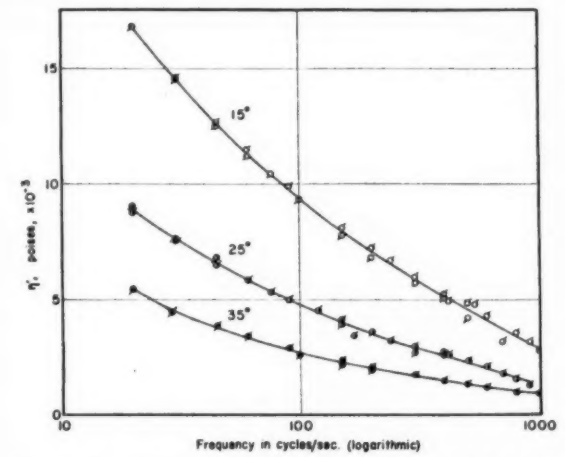

FIG. 6. Real part of dynamic viscosity of polyisobutylene, $\bar{M}_\eta = 1.2 \times 10^6$, plotted against the logarithm of the frequency. Key to tags same as in Fig. 5.

At each temperature and molecular weight, G' increases and η' decreases with increasing frequency. At any given frequency, both G' and η' decrease markedly with increasing temperature. An increase in molecular weight from 0.47 to 1.2 million increases G' by less than 20 percent, and has very little effect on η'.

DISCUSSION

There appear to be no other data in the literature on uncrosslinked rubbers with which the above results can be compared. It is of interest, however, to compare them qualitatively with similar data for vulcanized rubbers. Both G' and η' for polyisobutylene are found to be of the same order of magnitude as values for various cured rubbers (natural, GR-S, and butyl) at comparable frequency and temperature, derived from the measurements of Dillon, Prettyman, and Hall,[19] Nolle,[20] and Guth.[21] A quantitative comparison is difficult to make at present because chemical constitution has a considerable influence on the detailed behavior, though not apparently the order of magnitude, of G' and η'.

Dependence on Temperature

The variation of both G' and η' of polyisobutylene with temperature can be quantitatively explained by a single assumption, that stresses in the material relax by a variety of flow processes but the temperature de-

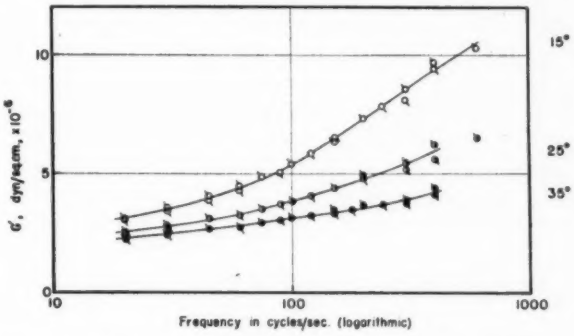

FIG. 7. Real part of dynamic rigidity of polyisobutylene, $\bar{M}_\eta = 0.47 \times 10^6$, plotted against the logarithm of the frequency. Upward tags, $2A/h = 24.5$; downward tags, $2A/h = 13.8$.

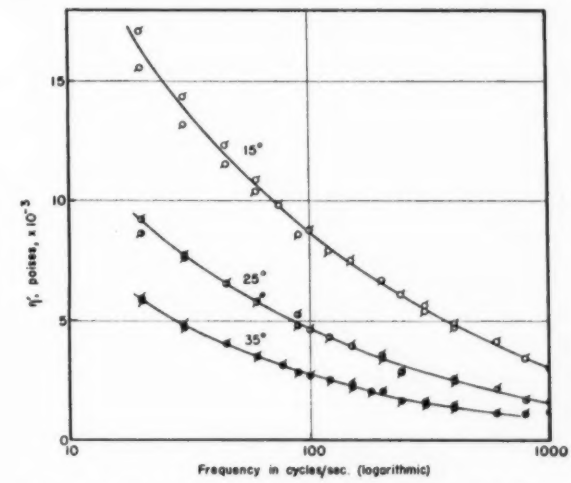

FIG. 8. Real part of dynamic viscosity of polyisobutylene, $\bar{M}_\eta = 0.47 \times 10^6$, plotted against the logarithm of the frequency. Keys to tags same as in Fig. 7.

[19] Dillon, Prettyman, and Hall, J. App. Phys. **15**, 309 (1944).
[20] A. W. Nolle, J. Polymer Sci. (to be published, and private communication); R. B. Blizard, Quarterly Progress Report of the Acoustics Laboratory, Massachusetts Institute of Technology (April-June, 1949).
[21] Ivey, Mrowca, and Guth, J. App. Phys. **20**, 486 (1949).

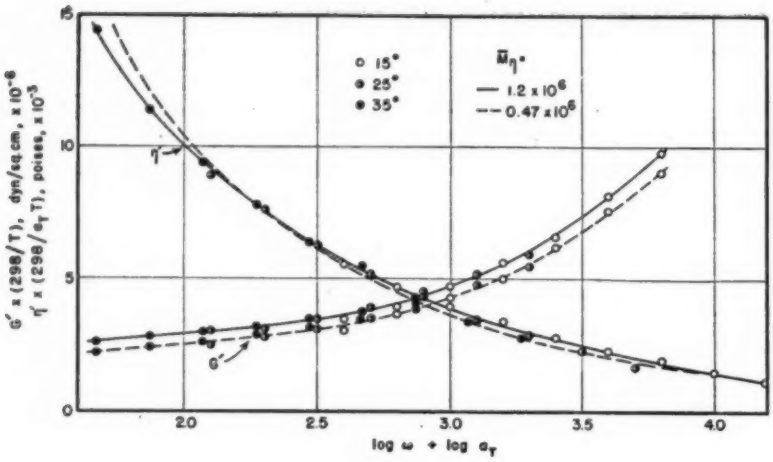

FIG. 9. Superposition of data at different temperatures: $G'(298/T)$ and $\eta'(298/a_T T)$ plotted against $\log\omega+\log a_T$. At 15°C, $a_T=3.16$; at 35°C, $a_T=0.38$. The points plotted for η' correspond to the solid curve ($\bar{M}_\eta=1.2\times10^6$); the points corresponding to the dashed curve ($\bar{M}_\eta=0.47\times10^6$) have been omitted to avoid confusion.

pendence of every such process is the same and identical with that of the steady flow viscosity. It is convenient to discuss the mechanical behavior in terms of an arbitrary distribution function of relaxation times, $G(\tau)$, such that the instantaneous rigidity is $\int_0^\infty G(\tau)d\tau$ and the steady flow viscosity is $\int_0^\infty \tau G(\tau)d\tau$. This has been employed by Andrews[3] and Kuhn,[22] and it is equivalent to the generalized Maxwell model of Alfrey and Doty;[23] that is, every spring $G(\tau)d\tau$ is relaxed by a dashpot $\tau G(\tau)d\tau$. In this case

$$G'=\int_0^\infty [G(\tau)\omega^2\tau^2/(1+\omega^2\tau^2)]d\tau$$

$$\text{and}\quad \eta'=\int_0^\infty [\tau G(\tau)/(1+\omega^2\tau^2)]d\tau.$$

Now, a change in temperature has two separate effects; it shifts the dispersion function to a different frequency region by altering the dimensionless variable $\omega\tau$, and it also changes the magnitudes of G' and η' by altering $G(\tau)$ and $\tau G(\tau)$, respectively. We will suppose that the rigidity mechanisms are of the nature of rubberlike elasticity and therefore that every $G(\tau)$ is proportional to the absolute temperature; this gives a slightly better fit with the data than does the assumption that the rigidity is independent of temperature, although the experimental temperature range is too small to test it critically. Then if a temperature change from T_0 to T (on the absolute scale) changes every relaxation time by a factor a_T, every contribution to G' will be multiplied by T/T_0 and shifted along a logarithmic frequency scale a distance $-\log a_T$; every contribution to η' will be multiplied by $a_T T/T_0$ and also shifted $-\log a_T$ along the $\log\omega$-axis. Moreover, the steady flow viscosity, η, will also be changed by the factor $a_T T/T_0$. Accordingly, choosing $T_0=298$°K, if we plot $G'(298/T)$ and $\eta'(298/a_T T)$ against $\log\omega+\log a_T$, data at all different temperatures should superpose.

Very successful superposition of both G' and η' for both molecular weights is obtained by selecting a_T values of 3.16 at 15°C and 0.37 at 35°C (Fig. 9). These values also fit in quite well with the temperature dependence of the steady flow viscosity of polyisobutylene given by Fox and Flory,[24] who showed that the apparent activation energy for viscous flow is independent of molecular weight and equal to 19.3 kcal. at 0° and 13.7 kcal. at 50°C. From the relation $\eta/\eta_{298}=a_T T/298$ we calculate 19 kcal. at 20° and 16.5 kcal. at 30°C. The values obtained by Andrews[3] from stress relaxation measurements are also from 15 to 16 kcal. in the range from 30° to 100°C. It appears therefore that all the flow processes in the relaxation spectrum which are reflected in these various measurements have the same temperature dependence.

Dependence on Molecular Weight

Figure 9 also emphasizes the very slight dependence of G' and η' on molecular weight. This indicates that the dynamic mechanical behavior involves primarily entanglements of molecular segments which are short compared with the total length, and that the loose ends of the molecules have relatively little significance. On this basis, it would be expected that η' would show still less dependence than G', as is indeed observed, because the contributions to η' come largely from mechanisms with short relaxation times (corresponding probably to shorter segments), while the contributions to G' come largely from mechanisms with long relaxation times (corresponding to long segments which could be perceptibly influenced by loose ends). By contrast, the steady flow viscosities of these two samples (if the viscosity average molecular weight can be assumed approximately proportional to the weight average) should differ by over a factor of ten,[24] and it is obvious from casual inspection of the samples that they do. This of course reflects the fact that the difficulty of moving entire molecules past each other increases markedly with molecular weight. Since η' must ap-

[22] Kuhn, Künzle, and Preissmann, Helv. Chim. Acta **30**, 307, 464 (1947).
[23] T. Alfrey and P. Doty, J. App. Phys. **16**, 700 (1945).
[24] T. G. Fox and P. J. Flory, J. Am. Chem. Soc. **70**, 2384 (1948).

proach η as $\omega \rightarrow 0$, the curves for η' must diverge widely at lower frequencies. Since η is of the order of 10^9 to 10^{10} poises,[24] and the maximum value of η' found here is about 10^4, the dispersion must extend well below the frequency range covered experimentally.

Dependence on Frequency

Figure 9 provides a picture of the dispersion functions for G' and η' over about two decades of frequency. There is at present no theoretical basis for interpretation of this dispersion. The relaxation distribution function derived by Kirkwood,[25] and the semi-empirical function of Kuhn,[22] are devised for crosslinked rubbers, and are not applicable to a linear polymer; tests have shown that neither fits our data for polyisobutylene. The "step" function used by Andrews[3] to describe stress relaxation in polyisobutylene provides (for frequencies not too close to the reciprocals of the cut-off times) that G' should be a linear function of $\log\omega$ and η' should be inversely proportional to ω; neither of these relations is even approximately fulfilled.

If a relaxation distribution function, either theoretical or empirical, can be obtained to fit our dynamic data, it should also apply to the transient stress relaxation data of Andrews,[3] with the reservation that his experiments involved extension rather than shear and represented a far greater extent of strain. However, a considerable gap in time scale remains to be bridged between the shortest transient measurement (about ten seconds) and the longest period in dynamic measurement (about one-twentieth of a second). Experimental data in this range are needed to provide a complete picture from which the mechanism of elasticity and elastic loss in uncrosslinked rubbers may be deduced.

[25] J. G. Kirkwood, J. Chem. Phys. **14**, 51 (1946).

ACKNOWLEDGMENT

This work was supported in part by a grant from Research Corporation, and in part by the Research Committee of the Graduate School of the University of Wisconsin from funds supplied by the Wisconsin Alumni Research Foundation. We are deeply indebted to Mr. Frederic W. Schremp for help in the detailed design of the apparatus, and to Mr. Thomas R. Lloyd for painstaking work in its construction. Professor F. T. Adler and Professor W. T. Thomson have given valuable advice in problems relating to mechanical resonances in the apparatus. We are grateful to Carbide and Carbon Chemicals Corporation for support through its Fellowship in Physical Chemistry.

The Dielectric Strength of Gaseous Fluorocarbons

W. A. Wilson, J. H. Simons, and T. J. Brice

Fluorine Laboratories, The Pennsylvania State College, State College, Pennsylvania

(Received August 16, 1949)

The sixty-cycle dielectric strengths of some gaseous fluorocarbons have been measured between three differently shaped electrode pairs at pressures up to three atmospheres. The breakdown potentials for propforane, butforane, and pentforane were found, in most instances, to be equal to or greater than those for sulfur hexafluoride under comparable conditions, and to be far greater than those for nitrogen. Fluorocarbons thus have possible uses as gaseous insulators in high voltage apparatus.

A CONSIDERABLE amount of research has been carried out during the past few years in evaluating the dielectric strengths of various gaseous materials. Rodine and Herb[1] found that if carbon tetrachloride vapors were added to air or nitrogen, the dielectric strength was raised considerably. Charlton and Cooper[2] measured the dielectric strength of numerous gases and found that the Freon compounds had much higher dielectric strengths than nitrogen under the same experimental conditions. Camilli and Chapman[3] have found that sulfur hexafluoride and dichloromethforane have dielectric strengths much higher than nitrogen, and, when the gases were under about three atmospheres pressure, the dielectric strengths of these compounds approached values obtained for transformer oils. These results have been utilized commercially by insulating certain high voltage condensers, transformers, and electrostatic generators with various compressed gases.

During this research, the sixty-cycle dielectric strengths of various fluorocarbon gases and vapors were measured. Propforane, butforane, and pentforane were found, in most instances, to have dielectric strengths equal to or greater than sulfur hexafluoride under similar conditions.

The experiments were carried out in three glass cells containing copper electrode pairs of different shapes. One cell contained flat, circular disks of 0.477-inch diameters, the second had hemispheres of 0.5-inch radii, and the third had a hemisphere of 0.5-inch radius and a pointed 0.149-inch rod as electrodes. The pointed

[1] M. D. Rodine and R. G. Herb, Phys. Rev. **51**, 508 (1937).
[2] E. E. Charlton and F. S. Cooper, Gen. Elec. Rev. **40**, 438 (1937).
[3] G. Camilli and J. J. Chapman, A.I.E.E. Technical Paper 47-240 (September, 1947).

end was a section of a right cone with 15°27′ semiangle and bases having diameters of 0.036 inch and 0.149 inch. The electrodes were sealed symmetrically about common axes into round bottom Pyrex flasks and subsequently sealed to a portable vacuum system. A 100 ml flask was used for the disk electrodes, and one liter flasks were used for the other cells. These electrode shapes are similar to those generally used for this type of measurement.[3] The pair of hemispheres gives an approximately uniform field, while the hemisphere-rod arrangement gives a very non-uniform field.

A small laboratory transformer was used for producing voltages less than 35 kv. For higher voltages, a transformer capable of producing 150 kv was used. The voltage measurements were made with an accuracy of ±5 percent.

The fluorocarbons were prepared by the electrochemical process described by Simons.[4] The samples used were fractions of narrow boiling range from distillations through efficient columns. Molecular weight values were within ±0.5 percent of the theoretical values. The samples were dried over P_2O_5 before distillation and again just prior to the dielectric strength measurements.

At the dielectric breakdown potential, arcing between the electrodes occurs. It was noticed that, except in the case of methforane, this arcing in fluorocarbons resulted in the formation of a thin layer of black material, probably carbon, on the surfaces of the electrodes and the glass flask; and that a pressure rise of about one per cent of the total pressure occurred. This latter observation seemed to indicate that lower molecular weight fluorocarbons formed to a slight extent during arcing. Since the last of a series of five breakdown potentials on a given gas sample was essentially the same as the first, it was evident that small amounts of these materials did not noticeably affect the dielectric strength of the fluorocarbon. In addition, the dielectric breakdown potentials of fresh gas samples were found to be the same, within experimental error, in a cell coated with the solid as in a new, uncoated cell.

The dielectric strengths are plotted against pressure on Figs. 1 to 5. Each plot represents the results obtained using a different electrode pair or a different electrode gap spacing. Sulfur hexafluoride and nitrogen were tested for comparison.

The tests using the disk electrodes (Fig. 1) were carried out at a gap spacing of 0.125 inch. The dielectric strength of the fluorocarbons increased with increasing molecular weight, and all the fluorocarbons had values higher than those for nitrogen. Butforane and pentforane had dielectric strengths higher than sulfur hexa-

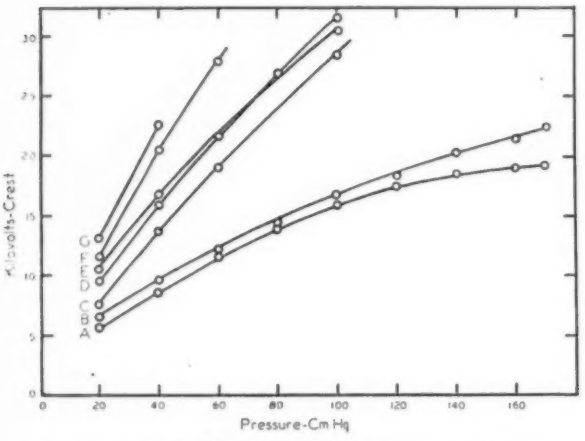

FIG. 1. 60-cycle dielectric strength between 0.477-inch diameter disks: Gap; 0.125 inch.

A—Nitrogen
B—Methforane
C—Ethforane
D—Propforane
E—Sulfur Hexafluoride
F—Butforane
G—Pentforane

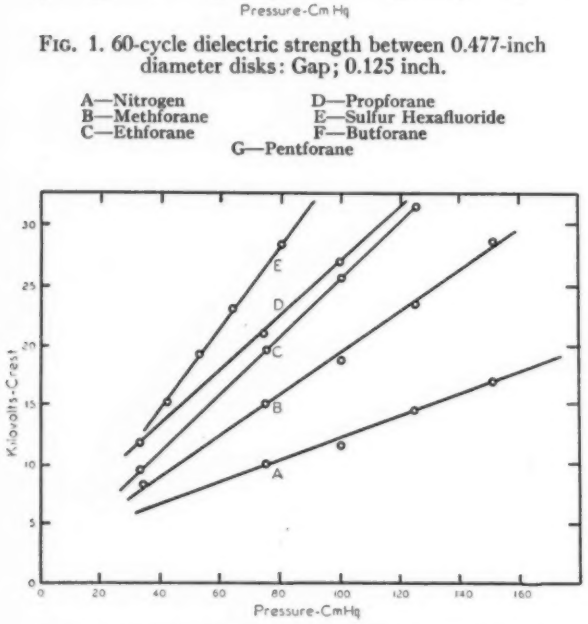

FIG. 2. 60-cycle dielectric strength between 0.500-inch radii hemispheres: Gap; 0.100 inch.

A—Nitrogen
B—Ethforane
C—Propforane
D—Sulfur Hexafluoride
E—Butforane

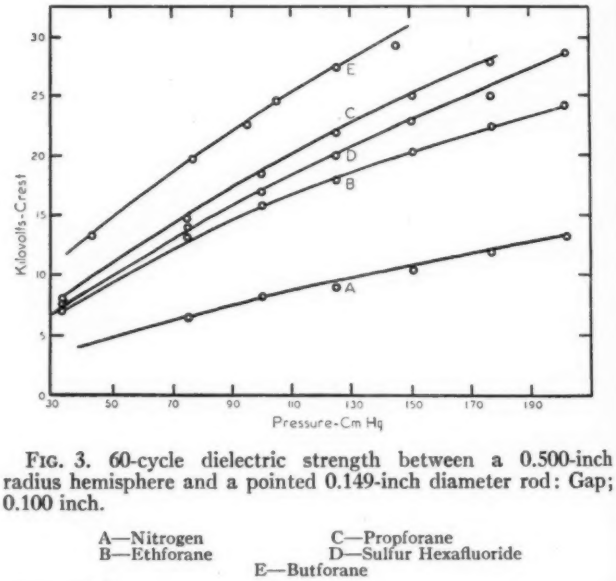

FIG. 3. 60-cycle dielectric strength between a 0.500-inch radius hemisphere and a pointed 0.149-inch diameter rod: Gap; 0.100 inch.

A—Nitrogen
B—Ethforane
C—Propforane
D—Sulfur Hexafluoride
E—Butforane

[4] J. H. Simons and co-workers, J. Electrochem. Soc. **95**, 47 (1949).

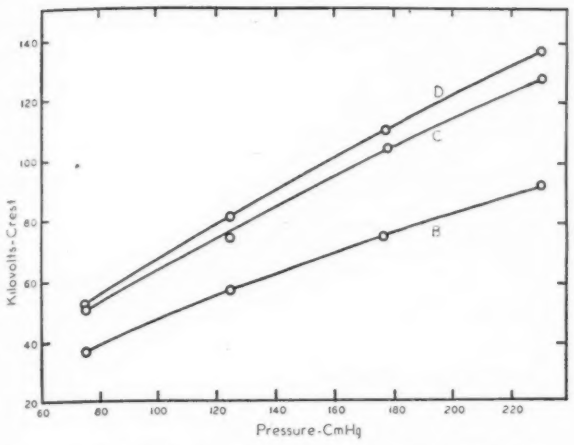

FIG. 4. 60-cycle dielectric strength between 0.500-inch radii hemispheres: Gap; 0.25 inch.

B—Ethforane
C—Propforane
D—Sulfur Hexafluoride

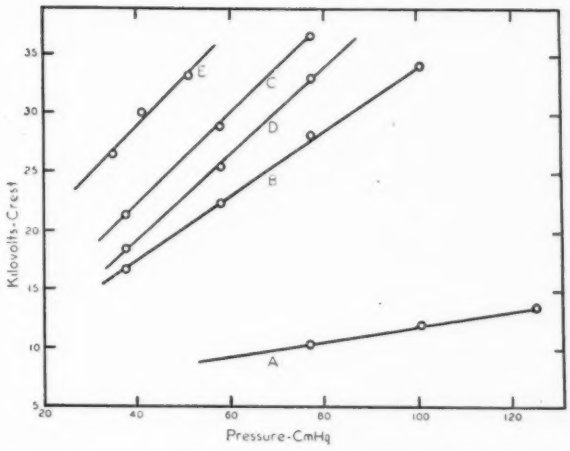

FIG. 5. 60-cycle dielectric strength between a 0.500-inch radius hemisphere and a pointed 0.149-inch diameter rod: Gap; 0.25 inch.

A—Nitrogen
B—Ethforane
C—Propforane
D—Sulfur Hexafluoride
E—Butforane

fluoride. The dielectric strength of propforane was lower than that of sulfur hexafluoride at pressures less than one atmosphere and higher than that of sulfur hexafluoride at pressures above one atmosphere.

The tests using the pair of hemispheres (Figs. 2 and 4) were carried out at gap spacings of 0.100 inch and 0.250 inch. The breakdown potentials obtained at equal pressures were roughly proportional to the gap spacing. The dielectric strength of propforane was somewhat lower than that of sulfur hexafluoride at all pressures used. The general pattern of relative dielectric strength values for the other compounds was the same as obtained with the pair of disks.

The tests using the hemisphere-point electrode pair (Figs. 3 and 5) were also carried out at gap spacing of 0.100 inch and 0.250 inch. As was expected, the breakdown potentials obtained when using this electrode pair were lower than those obtained when using the symmetrical electrode pairs which produced more uniform fields. When tested with this pair of electrodes, the dielectric strength of propforane was higher than that of sulfur hexafluoride at all pressures used. The pattern of relative values for the other compounds was the same as was obtained with the other electrode pairs.

An interesting correlation of the dielectric strengths of fluorocarbons is illustrated in Fig. 6. The dielectric strengths of fluorocarbons (taken from Fig. 1) are plotted against their gas densities. As a first approximation, the dielectric strength is proportional to the gas density, independent of the compound. Similar treatment of the data obtained when using the other electrode pairs gave similar results.

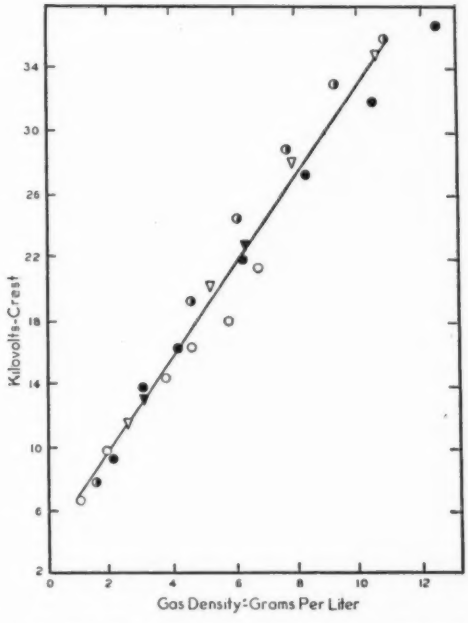

FIG. 6. 60-cycle dielectric strength between 0.477-inch diameter disks: Gap; 0.125 inch.

○ Methforane
◑ Ethforane
▲ Pentforane
● Propforane
△ Butforane

ACKNOWLEDGMENT

The authors wish to thank Professor R. E. Armington of the Electrical Engineering Department of The Pennsylvania State College for his assistance in obtaining the measurements reported here and the Electrical Engineering Department for allowing the use of their high voltage transformer.

This work was done under the sponsorship of the Minnesota Mining & Manufacturing Company, St. Paul, Minnesota.

Complex Stressing of Polyethylene

I. L. Hopkins, W. O. Baker, and J. B. Howard
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received August 17, 1949)

Polyethylene of such molecular weight and structure that it readily fibers or cold draws to 300–600 percent elongation by usual uniaxial, tensile stressing may react quite differently under biaxial tension. When biaxial tension in 1:1 ratio is applied to a diaphragm, some polymers show brittle fracture with <20 percent elongation at break. However, if the average molecular weight of such polyethylenes is shifted upward by crude fractionation, or an initially higher average is used, the polymers orient under complex stresses. Then, they usually elongate several hundred percent before rupture. Variations in crystallinity are also significant, although most technical polyethylene soon attains at room temperature enough crystallinity so that this factor does not cause big differences.

Although the whole study is so far preliminary, x-ray scattering of stressed samples suggests that preferred glide on certain crystallite planes tends to occur as the yield point approaches. These are such as to inhibit smooth alignment of the long chain axis in the direction of stressing. This could lead to brittleness.

Apparatus for complex stressing of sheets and tubes is described. Strains are taken from coordinates printed on the sample by the silk screen process. High speed stressing was also observed. The speed of retraction of amorphous polyethylene chains approaches that of rubber.

THE usual tensile strength or stress-strain measurement for plastics and rubbers involves uniaxial stretching of a long, thin strip. With such tension, polyethylene, and other microcrystalline polymers such as the polyamides and polyesters, give a steep initial stress *vs.* strain curve, up to 20–50 percent elongation. Then, "yield" occurs, and a quite flat curve delineates cold-drawing or fibering, marked by crystallite orientation. Finally, between 400 and 700 percent elongation, depending on the polymer, rate of stressing, degree of crystallinity, etc., the curve bends upward again and rupture soon occurs.[1,2]

However, two-directional or biaxial stress-strain properties have also been observed for rubber[3,4,5,6] and for thin Nylon films.[7] In these cases, while the form of the stress-strain curve is different than with uniaxial tension, the strains at rupture have apparently the usual values—of several hundred percent.

With polyethylene, we have found that biaxial stressing of some commercial grades (determined by some function of average molecular weight and chain linearity, such as melt viscosity), causes brittle fracture at 30 percent elongation or less. This is with a polymer having simple tensile properties of curves *C*—uniaxial in Fig. 1, wherein it breaks only after 300–500 percent elongation. Somewhat higher molecular weight polymer, however, shows much higher elongation at rupture, perhaps even approaching the lower uniaxial values. An example of this appears in curves *A* of Fig. 1. In these curves, the end point does *not* indicate rupture, as it does in curves *C*, except for the uniaxial tension curves *A*. Rather, on curves *A* for biaxial tension, the end of the curve denotes yield, where an oriented bubble begins to form on the stressed diaphragm. Accordingly, the material of curves *A* begins to orient at ~70 percent elongation, and draws down to 200 percent or more in the bubble before rupture. Nevertheless, its uniaxial curve is seen, in Fig. 1, to come quite close to that of material *C*.

The relative differences in average molecular weight of these two polymers are represented in Table I, by their inherent viscosities.

$$(\ln\eta_r)/C$$

in *m*-xylene (90 percent) at 85°C, with $C=0.5$ g/100-cc solution, and with η_r = time of flow of solution/time of flow of solvent.

The inherent viscosity values agree with the *uniaxial* stress-strain data of Fig. 1 in showing that both materials *A* and *C* are well within what is commonly called the "super-polymer" "cold drawing" range. Biaxial stresses thus have brought out a wholly unexpected and dramatic sensitivity of this sort of me-

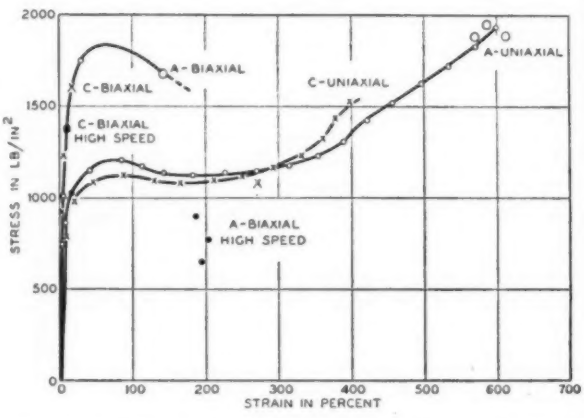

FIG. 1. Stress-strain curves of polyethylenes *A* and *C* under uniaxial and biaxial tension. (Large circles (for *A*) and crosses (for *C*) are rupture points, except that the large circle for *A*-biaxial is the yield point.)

[1] Hahn, Macht, and Fletcher, Ind. Eng. Chem. **37**, 526 (1945), show a fairly typical stress-strain curve for polyethylene.
[2] E. Hunter and W. G. Oakes, Brit. Plastics **17**, 94 (1945).
[3] P. Prache, Caoutchouc et gutta-percha **27**, 15, 125 (1930).
[4] J. R. Sheppard and W. J. Clapson, Ind. Eng. Chem. **24**, 782 (1932).
[5] C. F. Flint and W. J. S. Naunton, Trans. Inst. Rubber Ind. **12**, 367 (1937).
[6] L. R. G. Treloar, Trans. Faraday Soc. **40**, 59 (1944); Trans. Inst. Rubber Ind. **19**, 201 (1944).
[7] J. Miklowitz, J. Colloid Sci. **2**, 193, 217 (1947).

chanical behavior to average molecular weight as reflected in over-all molecular cohesion. Interestingly, the inherent viscosity for polymer C is actually higher than that for polymer A. The melt viscosities are reversed, however, with polymer A higher; it may be a more branched material although the over-all degreees of crystallinity of the two as judged by x-ray scattering, are similar. Further, polymer A has a definitely smaller fraction of low molecular weight, soluble in boiling chloroform, than polymer C; this is important, as judged from scores of similar comparisons made in this study.

For instance, polymer D in Table I showed, like polymer C, low elongation and brittle fracture under biaxial tension. It was then dissolved in pure trichloroethylene, and the system equilibrated at 43.8°C. Sixty-one percent by weight of the original sample was insoluble at this temperature; it had the properties of D' in Table I. The fraction of D' soluble in boiling chloroform was reduced fourfold from that of D, and the polymer was converted to high elongation cold-drawing behavior. However, its stress at yield was no higher than that of the unfractionated material (at rupture). By contrast, polymer D'' represents the 25 percent of polymer D insoluble in trichloroethylene between 43.8° and 25°C. Its density is a little higher than that of D'; its inherent viscosity is a little higher than that of the parent D; this may mean more chain linearity and crystallinity. But the portion soluble in boiling $CHCl_3$ has risen to 19 percent and the ultimate elongation under biaxial stress has fallen to 28 percent, accompanied by brittle fracture.

A', in Table I, reflects still another interesting aspect of biaxial mechanics. It is the fraction, 91.5 percent of polymer A, insoluble in the benzene-isopropanol azeotrope at its boiling point. The portion of A' soluble in boiling $CHCl_3$ is tiny, the density is still low, however, as well as inherent viscosity—some balance between removal of easily soluble species, but retention of (presumably) branched species appears. Hence, the material exhibits the orientation and tenacity of the parent A, but the elongation at yield *drops* sharply. This is a very typical behavior of "tough" polyethylene, and should not be confused with an approach to the low elongation at simultaneous yield and rupture of the polymers showing brittle fractures. In other words, as a function of average molecular weight and structure, the elongation at yield of polyethylene goes through a maximum, while the ultimate tenacity increases continuously.

TABLE I. Properties of polyethylene under biaxial tension.

Polymer	Density 25°C	$\ln\eta_r/C$	Percent soluble in boiling chloroform (N.T.P.)*	Strain at "yield" %	Strain at rupture %	Stress at "yield" p.s.i.
A	0.9191	0.933	9.6	70	~200	1600
A'(=91.5% of A)	0.9171	0.974	1.9	44	~200	1670
C	0.9197	1.007	14.5	17	17	1600
D	0.9167	0.879	12.6	24	24	1480
D'(=61% of D)	0.9191	1.007	3.8	66	~200	1450
D''(=25% of D)	0.9197	0.902	19.0	28	28	1360

* Amount soluble under empirical conditions.

This would have to be established finally by detailed fractionation of polyethylene; this has been explored, but not solved for reasons well recognized.[8] However, some additional examples of reaction of unfractionated, commercial-type polyethylenes to biaxial stress appear in Fig. 2. Apparently, with these molecular weight distributions, brittle fractures and low elongations do not occur much above an inherent viscosity of 1.0. Then, the above effect, where the still higher molecular weight polymers again show reduced elongations at yield, although they orient highly before rupture, appears at the extreme right in Fig. 2. Many additional points to those in Fig. 2 have been obtained, but they still do not allow construction of what might be a convex-upward correlation curve. A new correlation technique,[9] which has been especially useful in experimental studies of this kind, shows that extreme, or peripheral, association is probable in Fig. 2, however.

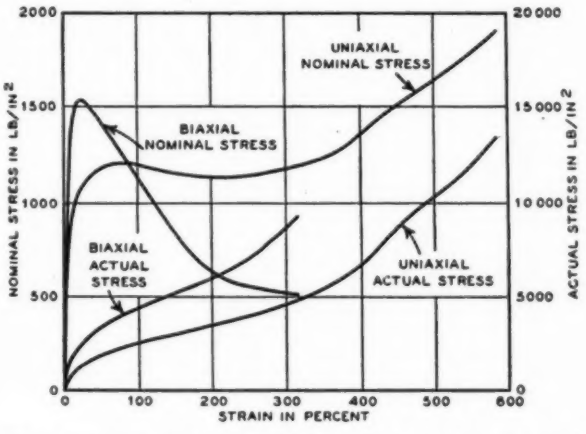

FIG. 1A. Stress-strain curves of polyethylene A under uniaxial tension, and biaxial tension with liquid medium. The biaxial nominal stress curve passes through and somewhat beyond the area of rupture points determined by the high speed tests with nitrogen as the pressure medium.

EXPERIMENTAL METHOD

Biaxial stresses were applied by N_2 gas pressure on a polymer disk. The cell in which the disks, in the range 0.080-in. to 0.140-in. thick, were clamped is shown in Fig. 3. The sample was illuminated by floodlights, behind heat-absorbing glass plates. It was photographed during straining, sometimes with a movie camera, but usually at intervals with the fixed camera shown behind the lights. Stress measurements were recorded by holding up a card bearing a definite figure for gas pressure (250 is shown in figure) and photographing the

[8] R. B. Richards, Trans. Faraday Soc. **42**, 10 (1946).
[9] P. S. Olmstead and J. W. Tukey, Ann. Math. Statistics **18**, 495 (1947).

card and the strained specimen just when the gas gauge reached the given figure.

The specimens were either clear polyethylene or a compound containing 1 percent or 2 percent by weight of fine carbon black, which had no effect on these physical properties. They were compression molded so as to be isotropic and free from included stresses. At first, 0.10-in. coordinate grids were photographed on them after coating with emulsion. However, it was found that the silk screen method of printing, long used for posters and textiles, worked very well with either a black or white ink containing a chlorinated rubber base. The ink had no effect of either crazing or swelling the

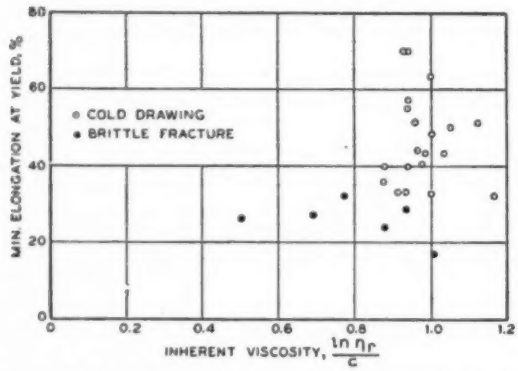

FIG. 2. Minimum elongation at yield under biaxial tension plotted against inherent viscosity of dilute solutions of polyethylene of various average molecular weights.

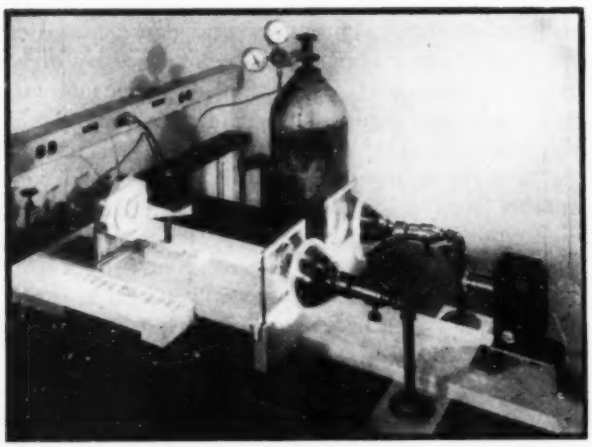

FIG. 3. Experimental arrangement for determining biaxial stress-strain curve. Diaphragm of polymer, bearing coordinates, appears below sign showing that it was subject at instant it was photographed (camera behind floodlights) to gauge pressure of 250 p.s.i.

polymer. These coordinates appear on the specimen shown in Fig. 3. Figure 4 shows typical drawn (upper and lower left) *vs.* brittle (lower right, with stick to mark crack) fractures. The white disk in Fig. 4 showed the reduced elongation at yield encountered for samples of high melt viscosity. Figure 3 shows a mirror obliquely placed along side the specimen holder so that the bulge of the diaphragm could be measured on the photograph along with the expansion of coordinates. From this, the stress in p.s.i. was computed as

$$PR/2t,$$

P being gas pressure in p.s.i., R, the radius of the (spherical) bulge and t, the thickness of the sheet in inches. Strain was obtained directly from the expansion of coordinates, which was the same in both directions (1:1 ratio) in most experiments. Some measurements were also done in a device in which cylindrical tubes of the polymer were blown up. This apparatus is shown in Fig. 5, with a polyethylene tube of 1 in. outside diameter and $\frac{1}{16}$-in. wall thickness. Coordinates were again printed directly on the specimen. Here, the principal stresses have a 2:1 ratio, but the results were quite comparable to the 1:1 ratio situation.

The rate of stressing was such that the pressure gauge rose uniformly over a period of 60 ± 10 sec. followed by abrupt rupture. Some experiments ten times as slow gave similar results and stress relaxation became significant for long periods; the present results suggest a study of stress relaxation under biaxial tension compared to the usual uniaxial case. Also, effects from very rapid straining were investigated. It was particularly desired here to be able to photograph the strains on the bubble which appeared as the high elongation samples oriented. This bubble is what blew into the crater shown on the upper sample of Fig. 4. Even with slow rates of over-all stressing with gas pressure, this bubble always appeared and burst so fast that ordinary photographs did not slow its growth. Hence, high speed

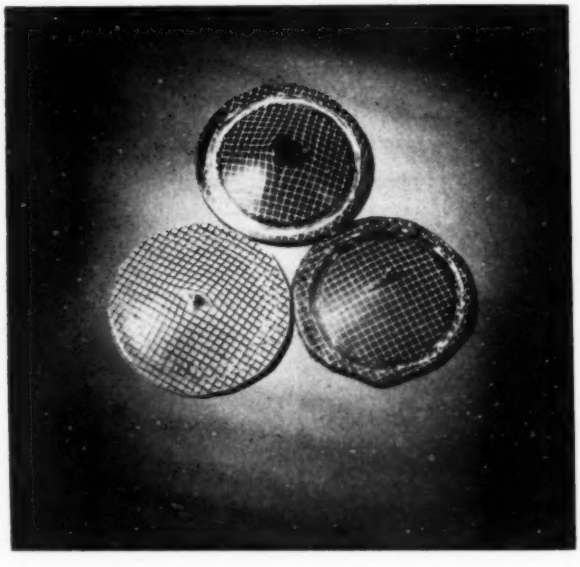

FIG. 4. Polyethylene disks ruptured by biaxial stress. Upper, moderate molecular weight, high yield strain, extensive cold drawing. Lower left, high molecular weight, lower yield strain, extensive cold drawing. Lower right, marginal molecular weight, low "yield" strain coinciding with brittle rupture, no cold drawing (stick is in crack to force it open for view).

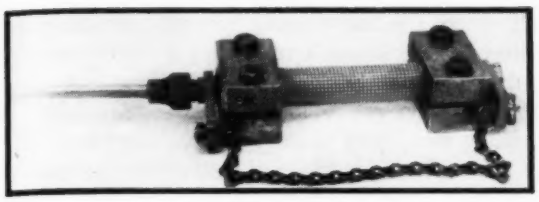

FIG. 5. Device for applying gas or hydraulic pressure to polymer tube bearing coordinates. Tube is clamped into fittings within steel blocks. (Here, rupture often occurs in position away from a single camera.)

photographs, with the Fastax camera, were made of specimens stressed and ruptured by sudden opening of the gas valve. The whole time from initial pressure to breaking was about 0.15 sec.; the bubble formed so rapidly that its life was roughly 10^{-3} sec. However, a few measurements of the coordinates on the bubble could be made, both as it formed and as the diaphragm relaxed after bursting. They are shown on Fig. 1 in terms of stress-strain assuming (1) that the stress-strain curve up to the point of yield (bubble formation) or brittle fracture is the same as in slower tests, (2) hence the pressures are the same at this point and in the case of the formation of a bubble do not change appreciably over the brief time interval required for the formation and bursting of the bubble, (3) that from the measured radius of the bubble and the assumed pressure, the stress in the bubble may be computed and (4) the strain in the bubble (ultimate elongation of the material) can be approximated by the over-all bubble dimensions. More appear on Fig. 6, where strains directly measured from the Fastax frames are plotted against time. The recovery from 130 percent back to about 50 percent total strain takes about 6 milliseconds. This is the same order velocity as for free contraction of soft rubber.[10,11] Probably amorphous chains of polyethylene are responsible; they are acting kinetic theory-wise compared to the usual sluggish "plastic" behavior of the polymer. In this connection, it will be noted later that biaxially stretched polyethylene seems to show an unusual orientation of the amorphous components (compared to their only slight orientation in an ordinary drawn fiber).

The high speed data still show the large relative differences in the polymers *A* and *C*. *C* breaks at 20 percent elongation, with brittle fracture, while *A* yields at a higher value, and cold draws to about 150–200 percent, as the points on the graph show, before breaking. Accordingly, failure of polymer *A* was not converted to brittle fracture by an "impact rupture" 10^4 times faster than the usual test. Incidentally, the growth of the brittle crack in polymer *C* could just about be seen in 10^{-4} sec. None of these data means, of course, that the apparent tenacity of the polymer is insensitive to the total *time* of stressing[12,13] (as in fatigue).

A modified apparatus is now being studied in which water (or aqueous anti-freeze solutions for low temperatures) is being used for pressure instead of gas. In this way, very little energy is stored during stressing, and the bubble formation for samples which orient, as well as the bulging of all samples, can be better controlled. Experiments done with water instead of nitrogen as the pressure transmitting medium bring test specimens

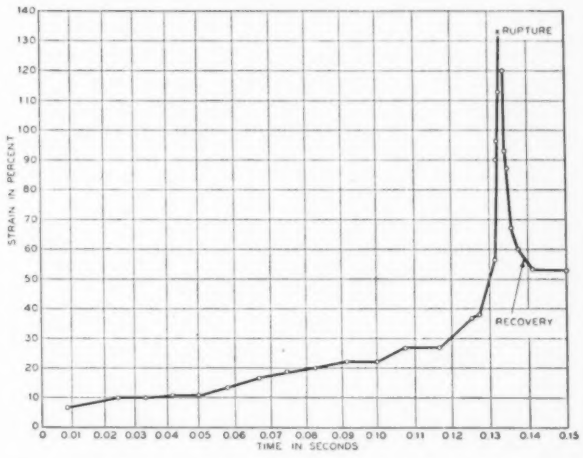

FIG. 6. Rate of straining of polyethylene diaphragm suddenly blown out by high gas pressure. The blown bubble itself remained highly oriented; recovery is for the rest of the surface.

gradually to the rupture point, rather than merely to the yield point, and show that biaxial strains of over 400 percent may be attained. Stress-strain curves for polyethylene *A* are shown in Fig. 1A. The biaxial nominal stress-strain curve of Fig. 1A passes through the biaxial high-speed rupture points of Fig. 1, confirming qualitatively the assumptions under which these points were calculated.

Reason for Biaxial Brittleness

Pending work in progress, and examination of other materials, only tentative qualitative explanations can be given for dependence of the elongation of polyethylene on stress complexity. In contrast to an assembly of long chain molecules like amorphous rubber, polyethylene at 25°C contains a major volume of small crystallites (100–300A in dimensions) and a minor volume of disordered matter. Hence, response of these crystallites to external stress, rather than gradual progressive orientation of individual chains as when rubber is stretched, is the critical process which will determine whether brittle fracture or cold drawing will occur. Of course, ultimately the long chain axes themselves line up in the direction of uniaxial fibering, as the

[10] R. B. Stambaugh, Phys. Rev. **65**, 250 (1944).
[11] Mrowca, Dart, and Guth, Phys. Rev. **66**, 32 (1944).
[12] R. N. Haward, Trans. Faraday Soc. **38**, 394 (1942); **39**, 267 (1943).
[13] T. Alfrey, *Mechanical Behavior of High Polymers* (Interscience Publishers, Inc., New York 1948), p. 481 ff.

crystallites are so oriented.[14] The mechanism of this adjustment,[15] the "shoulder" on cold drawn strips,[16] the dimensional changes on drawing, and stress-strain variations on redrawing[16] have also been considered in terms of crystallinity. But, less attention has been given to the striking structural changes in the early part of the uniaxial stress-strain curve, where it begins to flatten out on Fig. 1. Here, the long chain axes in the crystallites are actually put *at a large angle* ($\sim 70°$) *to the direction of external stress*[17,18] by the preferred glide on certain crystal planes. Thus, in the early stages of orientation, the x-ray reflections from 110 planes are split, up off the equator, while the other strong feature, the 200 reflections concentrate down on the equator.[17] At higher temperatures, where thermal agitation has disordered the crystallites[15,17] this anomalous orientation is absent,[18] and the long chain axes progressively align with the direction of stress.

Now, if such anisotropy of slip as this occurs during uniaxial pulling, it seems that under biaxial tension, a skewed orientation would result in which much of the pull between the chains was lateral. This would tend to pull them apart in the easiest way (since van der Waals bonds would support most of the stress). Hence, dependence of strength on molecular weight would be even more critical than in the uniaxial case.

Of course the "anomalous" orientation found at the beginning of uniaxial stressing of ordinary specimens (1 mm or more thick) may already reflect complex stressing. This is because of the appreciable shear attending stretching of such "thick" samples. An electron diffraction investigation of stretched polyethylene film initially about 1000A thick showed only progressive, normal, orientation of 110 planes,[19] but these films had some orientation during preparation. It would be interesting to see if such thin films, initially strictly isotropic, showed "anomalous" orientation from just stretching at 25°C.

In any case, x-ray scattering of polyethylene which has been sheared, or biaxially stretched (with some attendant shear), should show this competitive or "blocked" crystallite gliding. It does appear in the diagrams of Fig. 7. First, Fig. 7A shows the usual fiber diagram of highly, uniaxially, oriented polyethylene. The strong 110 and 200 reflections are well concentrated on the equator. The pattern (all are with crystal-monochromatized Cu $K\alpha$-radiation) also clearly exhibits the *unoriented*, amorphous components of the polymer.

Figure 7B is the diagram from a small sample roughly normal to the plane of principal shear in a tube which was twisted to failure. This tube was

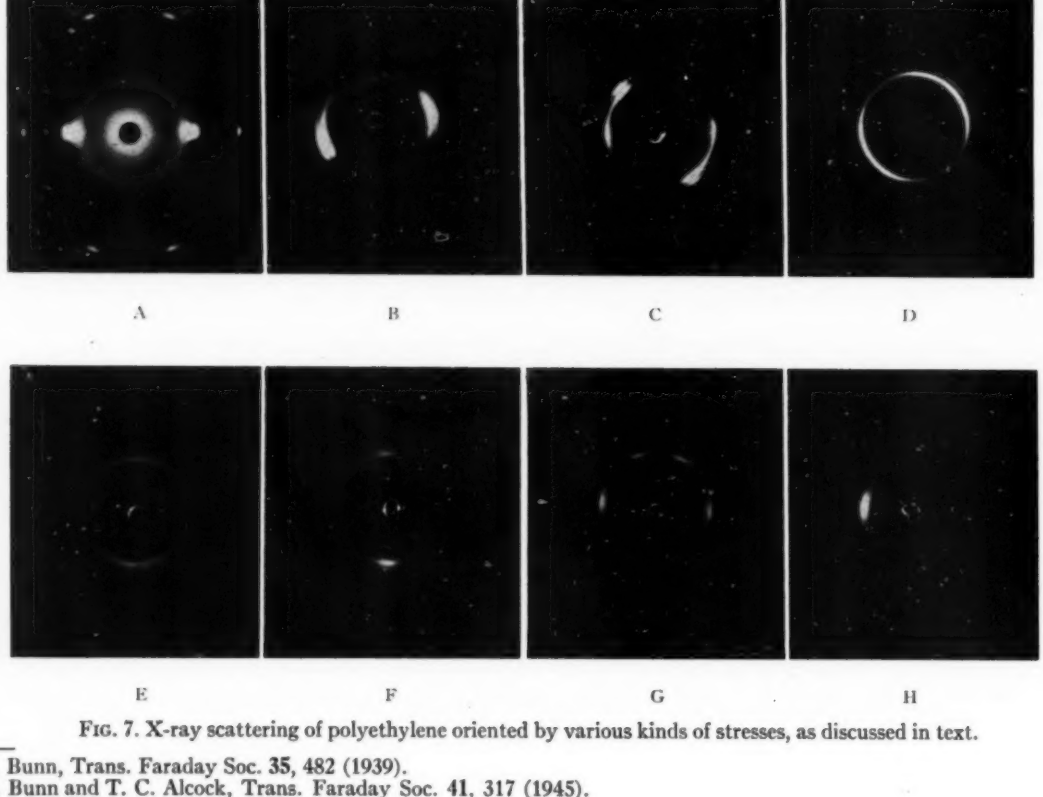

FIG. 7. X-ray scattering of polyethylene oriented by various kinds of stresses, as discussed in text.

[14] C. W. Bunn, Trans. Faraday Soc. **35**, 482 (1939).
[15] C. W. Bunn and T. C. Alcock, Trans. Faraday Soc. **41**, 317 (1945).
[16] W. M. D. Bryant, J. Polymer Sci. **2**, 547 (1947).
[17] W. O. Baker, *High Polymers* edited by Twiss (Reinhold Publishing Corporation, New York, 1945) p. 108.
[18] A. Brown, J. App. Phys. **20**, 552 (1949).
[19] A. Charlesby, Proc. Phys. Soc. **57**, 496 (1945).

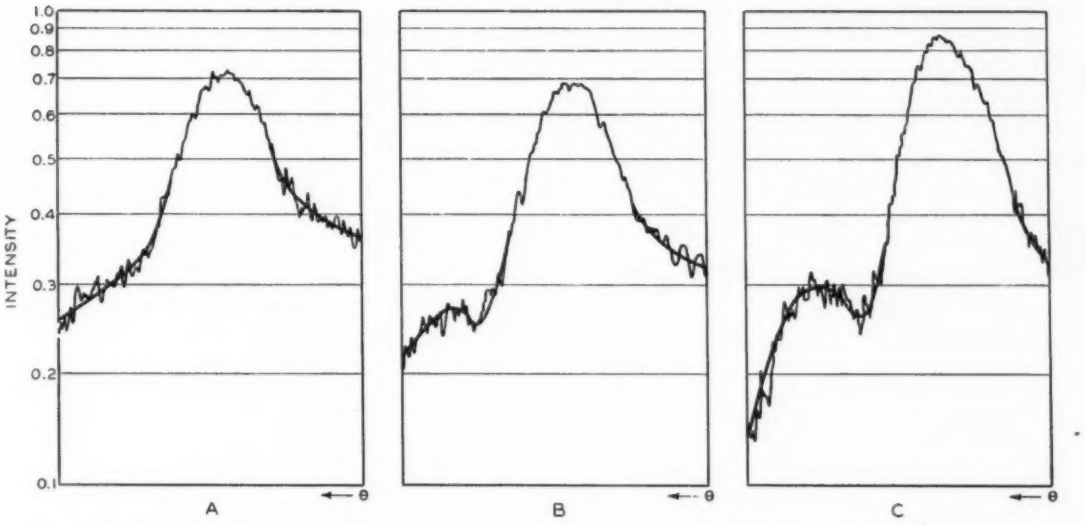

FIG. 8. Microphotometer tracings (uncorrected) showing x-ray scattering of principal features of unoriented polyethylene (A) immediately after quenching from melt to −78°C., (B) after same sample has stood 24 hr. at 25°C, (C) after annealing below melting point, and coming slowly at 25°C.

½ in. outside diameter, ¼ in. inside diameter, and about 12 in. long. It was twisted from the ends and sheared through near the middle. The B sample was at the rupture, and shows roughly normal alignment, although with a wide angular distribution and tapering of 110 and 200 intensities. Figure C is quite different, however. It shows what the polymer went through in getting to the final orientation just before rupture. It was obtained from a sample just back of the B sample. The 110 reflections are split, and also skewed with respect to the 200. The long chain axes by no means smoothly and gradually oriented in the direction of the principal shearing stress. However, Fig. 7D, whose sample came from a region of still lower shear than the C sample, shows that from the very outset these planes started to orient without "blocking."

Figures 7E–H exhibit these effects for specimens adjacent to the rupture of biaxially stressed disks which cold-drew or "bubbled," as shown at the top of Fig. 4. Samples were cut from the lips of the large circular bubble like that in Fig. 4. They were cut as little strips whose length lay along a radius of the test disk. Accordingly, they were somewhat tapered in thickness because of the drawing down at the rupture. When mounted with long axis vertical (like the fiber diagram of Fig. 7A), and the plane of the strip (and thus disk) normal to the x-ray beam, patterns like Fig. 7E result. From this point of view, the chain axes lie as though they had tried to form rings around the pole of the inflated disk. Figure 7F shows a similar diagram for a sample from a higher molecular weight polymer, such as in the white disk in Fig. 4. Here the stress at rupture was about 2000 p.s.i., as compared to 1500–1600 p.s.i. for the black, drawn disk. The orientation was more complete, but otherwise similar. When these radial strips were turned 90° around their length, so that the x-ray beam entered parallel to the plane of the strip (and the disk), diagrams like Fig. 7G were found. These indicate the "two-way" or cross-hatched orientation which has occurred in the plane. This is presumably because the crystallites have attempted to glide in the directions of principal stresses, but have been blocked because of the gliding anisotropy noted before. This results, again, in much of the external tensions being *across* chain axis and unless the molecular weight is quite high, remarkable brittle, unoriented fractures such as those of the third disk in Fig. 4 occur. The off-equator spacings in Fig. 7G are the familiar 4.17A and 3.70A of the *n*-paraffin lattice. The equatorial spacings are, however, the strong 4.17A (110) and a fuzzy fea-

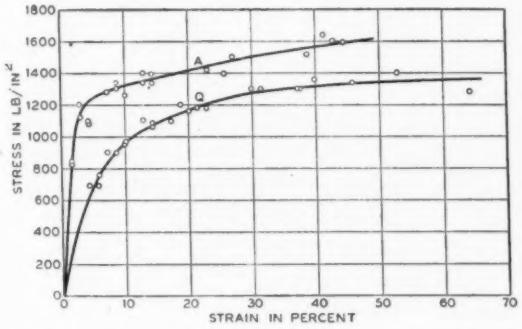

FIG. 9. Stress-strain curves of annealed and quenched polyethylene.

ture which has a Bragg spacing of 4.6A—presumably *oriented amorphous* chains whose unoriented halo scattering appears in Fig. 7A. This stretching out of amorphous chains may account for the quick, rubber-like contraction noted in connection with Fig. 6. Figure 7H shows the orientation appearing when the x-ray beam is parallel to the plane of and directed along the radius of the disk at the rupture.

DEGREE OF CRYSTALLINITY AND COMPLEX STRESSING

If preferred crystallite glide is so significant in reaction of polyethylene to complex stressing, large variations in crystallinity should be reflected in stress-strain properties. Degree of crystallinity of essentially simple chain polymers may be influenced in two ways. First, equilibrium crystallinity may diminish (or transform) with increase in temperature.[15, 17, 20, 21] Secondly, *metastable* reduction in crystallinity occurs in quenched or shock-cooled samples.[17, 22] The latter condition is particularly important as it arises, often in uncontrolled ways, in technical processing of thermoplastics. However, while the life of these metastable states of order may be very long (years) for polar polymers like the polyamides and cellulose derivatives it is short for weakly-interacting chains like polyethylene. Thus, while spherulites may not be able to grow, and hence quenched samples of polyethylene may stay quite transparent, lateral ordering of the chains occurs rapidly at 25°C.

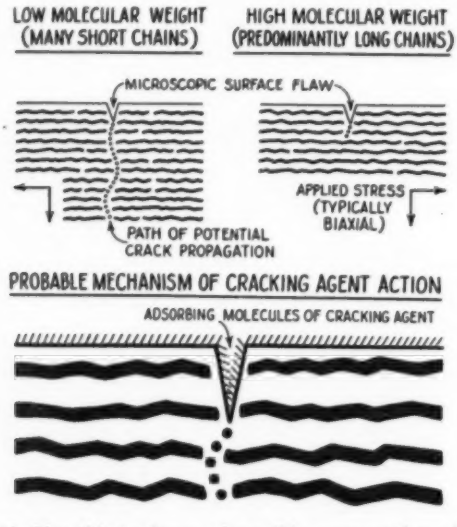

FIG. 10. Simplified scheme of possible cross section of sheet of biaxially stressed polyethylene (in which one of the stress arrows should be imagined as perpendicular to the page). Crack propagation is shown by dotted lines; electron micrographs of crack replicas indicate that paths may run around crystallites or spherulites. Cross-sectional size of cracking agent molecules should be considered as about same as that of polymer molecules. Replicas have shown that tiny surface flaws are present.

This is shown by the microphotometer tracings of x-ray diagrams, in Fig. 8. Specimen A was thoroughly quenched, by much more rapid cooling than could occur in practice. The 200 spacing is hardly resolved, and the 110 is diffuse with a half-breadth of 8.7 (in arbitrary units; the half-breadths were obtained after subtracting background corrections). While this sample had a lower yield stress than the same polymer annealed to condition C, its ultimate stress-strain properties were not radically different. Further, after 24 hr. at 25°C, the sample had re-ordered to give the pattern represented by Fig. 8B. The 200 spacing feature is resolved, although weak with a half-breadth of 7.4, while the 110 has intensified and sharpened to 7.7. This change indicates (1) that important physical differences in polyethylenes which have been processed with different thermal history will be lost after a few days standing at room temperature (this refers to lateral ordering, *not* to any long chain axis orientation or "molding strains"); (2) the energy differences between the metastable condition A and more ordered condition B are small, since activation energy for extensive transformation seems to be provided by kT at 25°C. This agrees with the small difference in physical properties between the two states, in contrast to the polar polymer case.[16]

Finally, the curve of Fig. 8C is for a sample which was allowed to solidify slowly, and was held at an elevated temperature for several hours. The 110 feature now has a half-breadth of 6.9, and the 200 of 7.4. This sample had maximum growth of spherulites, as well as presumably roughly[20] equilibrium crystallinity, but the conclusion is that over-all molecular order is not notably more than would eventually be reached by a practical sample standing at room temperature. Also, it was found that a polymer of molecular weight properties in the low elongation-brittle fracture class of Figs. 1 and 2 was not converted to a tougher, cold-drawing behavior by extreme quenching, and testing after standing a day. No doubt a polyethylene just on the verge of being able to orient under complex stressing, in the annealed condition could be oriented after thorough quenching. But these circumstances do not often arise; curves 8A and 8C represent practically the whole range of metastable states of the present type of commercial polyethylene. Detailed studies of polymers covering the 8A–8C extremes showed biaxial stress-strain curves of the latter lying about 240 p.s.i. above the disordered samples. Moduli from the first, linear portions of the curves were 13,200 p.s.i. for the quenched and 34,800 p.s.i. for the annealed polymers. The complete stress-strain curves appear in Fig. 9.

Stable reductions in degree of crystallinity at elevated temperatures are currently being studied in relation to complex stressing at those temperatures. Presumably near 100°C there will be easier and more isotropic crystallite orientation than at 24°C.[18] The over-all reduction in crystallites also shows up in stress-strain properties.[23] Hence, many of the effects in this report may be changed at other temperatures.

CRACKING AGENTS AND COMPLEX STRESSING

Crazing or micro-cracking of polymers has been long studied, especially in technical development of rigid

[20] R. B. Richards, Trans. Faraday Soc. **41**, 127 (1945).
[21] Raine, Richards, and Ryder, Trans. Faraday Soc. **41**, 56 (1945).
[22] Fuller, Baker, and Pape, J. Am. Chem. Soc. **62**, 3275 (1940).
[23] Stein, Krimm, and Tobolsky, Textile Research J. **19**, 8 (1949).

polymers such as polystyrene and polymethyl methacrylate, and in many finishes. It arises from shrinkage stresses when volatile components (such as unreacted monomers) escape, from external stresses,[24] and from internal stresses reflecting inhomogeneous molecular packing.[25] Spectacular acceleration of crazing occurs with application of many organic liquids or their vapors to the surface of these polymers, under either internal or external stress.[24,25] Stressed polyethylene (of certain molecular weight range) also shows this brittle cracking, by "mobile polar liquids."[8] The present work has revealed that complex stresses are strikingly effective in this surface-agent attack; biaxial tension causes extensive cracking in the presence of alcohols, esters, organic salts, etc., many times faster than uniaxial tension. However, as with the effects of biaxial tension alone the effects with surface-agents are also sensitively affected by molecular weight.

Cracking liquids have been called "surface-agents" above, following a mechanism postulated for crazing action, in connection with earlier work.[25] This is shown schematically in Fig. 10, where the "lyophilic" or non-polar parts of the adsorbed molecules of cracking liquid are assumed to penetrate the outermost layers of polyethylene chains. These adsorbed molecules presumably tend to spread to form Langmuir-type films. As they cover the polymer surface, they run along into tiny flaws and cracks and, at the apex of these cracks, probably exert a film spreading pressure. This film pressure, along with weakening of the polymer mole cohesion near the apex by swelling by the non-polar portions of the sorbed molecule, may add to an external stress to yield crack propagation. Biaxial external stresses, whatever the reason, cause especially rapid cracking in the presence of such agents.

Likewise, solvents like benzene, toluene, etc., which can actually swell polyethylene, even at room temperature, cause brittle fracture under biaxial stress. This seems to be just a weakening of the intrinsic structure, however, perhaps because of the drastic lowering of number average molecular weight in such a "plasticized" system. (Before complete penetration, these solvents presumably act as surface agents.)

[24] W. B. Klemperer, Applied Mechanics, T. von Kármán Anniversary Volume, 328 (1941).
[25] W. O. Baker, U. S. Patent 2,373,093 (1945).

GENERAL SIGNIFICANCE

Reactions of other microcrystalline polymers to complex stresses is being explored. Many queer cases of brittle failures in use by polymers normally considered (by uniaxial criteria) to be tough may have come from biaxial stressing in service. Certain brittle point tests (the test for incidence of brittle fracture, usually in rubbery materials at low temperatures) also seem to introduce varying amounts of biaxial tension in the bent sample. This may be one of the poorly controlled elements in brittle point evaluation. Likewise the great sensitivity to cracking agents of polymers already studied under complex stresses suggests that effects of liquid immersion baths in brittle tests may have been underestimated.

Bending of cable sheaths causes mostly biaxial stressing as do many common deformations of tubes, bottles, and the like. Likewise, paints and lacquers on tubular or other surfaces which are bent can become subject to intense biaxial stressing. In these cases, the strains are often small enough so that the given polymer material is commonly thought to have much reserve elongation before rupture, *again based on conventional uniaxial tests.* The present work at least indicates that many of these usual engineering practices must be re-examined. In the case of cable sheath, a complete upward revision of ideas of acceptable molecular weight polymer was necessary because of effects of complex stressing when the cable was bent and twisted.

Frozen, crystallized natural rubber is being studied to see if rubber crystallites not *formed* in the direction of stress show blocking. This would be a variation of the normal behavior found for initially amorphous rubber biaxially extended up through its presumed stress-crystallization point.[6]

Several technical tests such as paper bursting tests operate under complex stresses. Microcrystalline polymer films, especially those of interest because of low moisture permeability, such as polyethylene, rubber hydrochloride, polytetrafluorethylene, and polytrifluoro-monochlorethylene may show special properties in these tests.

ACKNOWLEDGMENTS

Mr. N. R. Pape has contributed broadly to these x-ray and mechanical studies. We wish also to thank Dr. R. D. Mindlin and Mr. W. T. Read for discussions about the diaphragm and tube tests.

D.C. Characteristics of Silicon and Germanium Point Contact Crystal Rectifiers. Part I. Experimental*

H. J. YEARIAN
Physics Department, Purdue University, Lafayette, Indiana
(Received August 29, 1949)

Typical d.c. current-voltage characteristics obtained for Si and Ge crystal rectifiers are described. A survey of the published theories of the rectifier shows that none of them will account for the principal features of the observed characteristics. The most obvious discrepancy is in the low resistance direction of flow where the logarithm of the current rises with increasing voltage at only a fraction of the rate indicated by the theories. The need for a more flexible theory is pointed out and the conditions which it must meet, both as a function of voltage and temperature are outlined.

THE development of microwave radar has been accompanied by wide use of point contact crystal rectifiers, and a revived interest in the details of their behavior. Since a complete understanding of their operation at high frequencies must be predicated upon their low frequency properties, extensive investigations of their d.c. characteristics have been made.

It is the purpose of Part I to give typical d.c. characteristic data of silicon and germanium rectifiers, the two types of interest at present in the microwave region,** and to show that the published theories of the rectifier do not account for the important features of these characteristics. Part II† will discuss an extension of the diode theory, the so-called multicontact theory, which is in agreement with the experiments.

The usual rectifier for high frequency applications consists of a metal point of approximately 10μ-diameter, in contact with a suitably alloyed crystal of silicon or germanium semi-conductor. The ratio of current at a given applied voltage in the low resistance, or forward, direction to that in the high resistance, or backward, direction, may be several hundred. The silicon now in use is a hole, or P type, conductor for which the direction of easy flow of electrons across the rectifying junction is from metal to semi-conductor. The polarity is opposite for germanium, since this material is used as an electron, or N type, conductor.

In series with the variable resistance of the rectifying barrier is the resistance encountered as the current spreads outward from the contact into the semi-conductor. This "spreading resistance" has a value, $R_s=\rho/(2d)$, where ρ is the resistivity of the semi-conductor and d is the diameter of the contact. All experimental data must be corrected for the potential drop in this resistance before comparison with theories of the rectifying barrier.

A number of theories of the rectifying contact have been developed[1] which differ in their treatment of the problem according to the assumptions made concerning the nature of the rectifying barrier or blocking layer.

The barrier may be assumed to be an artificial insulating layer on the surface of the semi-conductor, or a depletion layer formed near the surface by absence of the impurity centers which produce the conductivity of the semi-conductor. It may also be assumed to be the natural barrier formed when direct contact is made between a semi-conductor and a metal having a different work function.

The electron potential energy diagrams for these cases are shown schematically in Fig. 1, for an N type semi-conductor. Analogous diagrams may be made for the P type.

For the usual densities of donator states slightly below the bottom of the conduction band of the semi-conductor, the Fermi level lies approximately half way between the donator levels and the bottom of the conduction band. Well within the semi-conductor, when at equilibrium with no voltage applied, this level must lie at the same height as the top of the Fermi lake in the metal. When, therefore, the work function of the semi-conductor is less than that of the metal,†† a potential barrier is formed of height equal to the difference of the work functions, the contact potential, V_c.‡ With an insulating layer on the surface of the semi-

* This work was performed under OSRD Contract OEMsr-362 and was reported in an NDRC Report dated December 3, 1942, PB5189.

** The "High Back Voltage" type of germanium rectifiers have high frequency properties which are significantly different from the microwave type, but the low voltage d.c. characteristics are similar and most of the conclusions of this paper will apply. See reference 5, and NDRC Reports Purdue University: 14-341, PB5201, Nov. 1, 1944; 14-375, PB5202, Dec. 26, 1944; 14-413, PB5204, March 19, 1945.

† Johnson, Smith, and Yearian (to be published).

[1] W. Schottky, Zeits. f. Physik **118**, 539 (1941); N. F. Mott and R. W. Gurney *Electronic Processes in Ionic Crystals* (Oxford University Press, New York, 1940); W. Schottky and E. Spenke, Ver. Siemen-Werk. XVIII 31-68 August 1939; W. Schottky, Zeits. f. Physik **113**, 367 (1939); N. F. Mott, Proc. Roy. Soc. **A117** 27 (1939); Proc. Camb. Phil. Soc. **34**, 568 (1938); B. Davydow, Tech. Phys. U.S.S.R. **5**, 87 (1938); A. H. Wilson, Proc. Roy. Soc. **A136**, 487 (1932); J. Frenkel and A. Joffe, Physik. Zeits. Sowjetunion **1**, 60 (1932), Lothar Nordheim, Zeits. f. Physik **75**, 434 (1932); C. Wagner, Physik. Zeits. **32**, 641 (1931).

†† These are the work functions under existing conditions, not necessarily those for clean surfaces.

‡ J. Bardeen has investigated the effect of energy states on the surface of the semi-conductor (Surface States and Rectification at a Metal Semi-Conductor Contact, Phys. Rev. **71**, 717 (1947)) and has shown that the barrier height may be nearly independent of the metal work function. No major changes in the results of the rectifier theories outlined below will be occasioned by introduction of the surface states if "V_c" and "contact potential" are interpreted as meaning the height of the potential barrier, however formed.

conductor, either due to an artificial layer of an insulator or to a depletion of donators, there results a nearly linear variation of potential across this region. With no insulator nor depleted layer, when contact is made electrons near the surface will go over to the metal, leaving a positive space charge due to ionized donators. The potential fall in this case is quadratic and the barrier has a thickness, D, of approximately

$$D=[\epsilon V_c/(2\pi eN^+)]^{\frac{1}{2}}, \quad (1)$$

where ϵ is the dielectric constant and N^+ the density of ionized donators.

The theories may be further differentiated according to the assumption of a thin, or a relatively thick barrier layer.

For the case of an insulator sufficiently thin that electrons may tunnel through, Wilson,[1] and also Mott[1] have shown that if the current and the voltage across the contact are taken positive for electrons flowing from semi-conductor to metal, the net current density is

$$j=-4\pi me(kT)^2ph^{-3} \times\exp(-V_c/kT)[\exp(-eV/kT)-1], \quad (2)$$

where p is the transmission coefficient of the barrier, V is the voltage across the barrier (applied voltage V_a minus the potential drop in the spreading resistance) and the other constants have their usual significance.

As has been pointed out by Mott, this result is contrary to experiment since for an N type semi-conductor it gives the low resistance direction as corresponding to the flow of electrons from metal to SC $(-V)$ whereas, experimentally, this is the high resistance direction. Tunneling effects cannot, therefore, be the fundamental source of the rectification. Such effects are of importance, however, in the detailed analysis of the characteristic in the backward direction and will be discussed below.

In the case of a depletion layer for which the thickness, D, is large compared with the electron mean free path, so that conditions of diffusion exist in the barrier region, Mott[1] finds for the current density

$$j=ebn_0(V_c-V)D^{-1}[\exp(eV/kT)-1], \quad (3)$$

where b is the electron mobility in the semi-conductor and n_0 the electron density at the boundary of the semi-conductor in the absence of current.

In the forward direction for $0<V<V_c$, Eq. (3) is similar in behavior to that for tunneling, but the sign is reversed as experiment requires. The exponential is the dominant term, so that the slope of the curve ln$i$ *versus* V, $V\gg kT/e$, should be only slightly less than e/kT. In the backward direction, $V<0<V_c$, the current should quickly approach an approximately ohmic behavior of differential resistance $D/(ebn_0)$ determined by the first factor. The back current asymptote extrapolated to $V=0$ should give an intercept ebn_0V_c/D, the current in each direction across the barrier at zero applied voltage.

The effect of image force is to lower the barrier and this will have the effect of making the backward current increase more rapidly than indicated by Eqs. (2) or (3).

The theory of the thick natural barrier has been considered in detail by Schottky and collaborators[1] especially for the low voltage region. For semi-conductors of the "exhausted" type, in which the donators are completely ionized, the characteristic may be written in the form

$$j=\sigma_b(8\pi eN/\epsilon)^{\frac{1}{2}}(V_c-V)^{\frac{1}{2}}[\exp(eV/kT)-1] \times\{1-\exp[-e(V_c-V)/kT]\}^{-1}, \quad (4)$$

where N is the density of (completely) ionized donators and σ_b is the conductivity of the semi-conductor at the boundary. For $V_c\gg kT/e$ and $V\ll V_c$, the denominator is approximately unity. The square root varies only slowly, so that in the forward direction, V positive, the current increase is nearly exponential with a logarithmic slope of $\approx e/kT$. In the backward direction

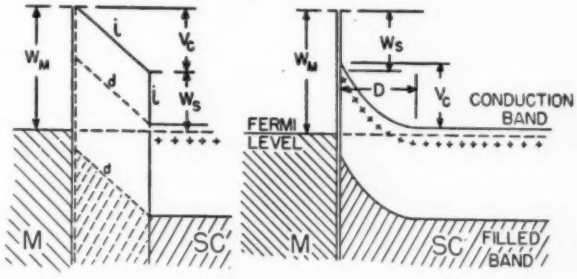

FIG. 1. Potential barrier between metal, M, and N type semi-conductor, SC, at equilibrium with no voltage applied. W_m and W_s, work functions of metal and semi-conductor; D, thickness of natural barrier; $V_c=W_m-W_s$, barrier height; $+++$ donator levels. Left, insulator (full lines i) between M and SC with potential differences V_c across insulating barrier; left (dotted lines d) depleted layer of SC next metal with potential difference V_c across this layer; right, natural barrier of height V_c. Levels at right of each figure will be moved up or down by amount of positive or negative voltage, respectively, applied across the contact.

the current should increase slowly according to the square root.

When the donators are not completely ionized, the so-called reserve region, Schottky and Spenke[1] have shown that at low voltages the factor corresponding to the logarithmic slope e/kT becomes $e/(\delta kT)$, where δ may vary up to three as a maximum. They also show that as the higher forward voltage region is approached, the characteristic curves for the reserve and exhausted cases merge. The logarithmic slope as measured at voltages above 0.1–0.2 volt, therefore, should be $\approx e/kT$ in either case. It is likewise shown that in this region there is no essential difference between the theories for depletion and non-depletion of donators. This is in agreement with the similarity of Eqs. (3) and (4) in the forward direction.

If the natural barrier is thin compared with the electron mean free path in the semi-conductor an electron makes no collisions in the barrier region, and the problem may be treated as that of a diode. This theory has

TABLE I. Electrical properties, mean free path and barrier thickness in several samples of germanium and silicon.

	Resistivity ohm cm	Hall coefficient cm^3 $climb^{-1}$	n $10^{13}cm^{-3}$	N $10^{13}cm^{-3}$	x_0 $10^{-7}cm$	D $10^{-7}cm$	l $10^{-7}cm$	ΔE e.v.
N type Ge	0.0665	280	26.3	26.3	41.5	141	273	0.051
	0.044	61	121	125	19.5	64.5	86	0.039
	0.0194	19.8	372	374	11.1	37.8	64	0.017
	0.00715	7.4	983	983	6.75	23.0	65	*
	0.00500	3.43	2150	2170	4.60	15.7	43	*
	0.0044	2.93	2510	2630	4.23	14.4	41	*
	0.079	104	70.9	76.8	25.2	85.7	82	0.047
	0.0535	77	95.8	105	21.9	74.5	79	0.072
	0.026	31	238	300	13.9	47.3	74	0.037
	0.0105	12.8	575	624	8.88	30.2	51	0.032
	0.0104	10.4	709	792	8.02	27.3	62	0.030
	0.0101	9.8	784	819	7.78	26.5	61	0.020
	0.0075	7.9	934	980	7.00	23.8	65	0.024
P type Si	0.048	5.4	1370	3400	9.0	30.6	7.9	0.081
	0.048	4.2	1750	4400	7.9	26.9	6.2	0.077
	0.029	1.74	4230	22,200	5.1	17.3	4.2	0.084

* Hall curve too flat to permit reliable calculation of ΔE.

been developed by Bethe[2] and Sachs[3] and gives the result

$$j = j_0 \exp(-eV_c/kT)[\exp(eV/kT)-1], \qquad (5)$$

where j_0 is the thermal current density in the semi-conductor, $j_0 = ne[kT/2\pi m)]^{\frac{1}{2}}$ and n is the electron density in the conduction band. The terms of Eq. (5) outside the brackets represent the electron current over the top of the barrier from metal to semi-conductor and with no voltage applied, the equal current in the opposite direction. As the voltage is varied the first current is constant (without image force or tunneling effects) while that from semi-conductor to metal varies exponentially with voltage. This theory therefore also gives a logarithmic slope in the forward direction of e/kT, while the current density in the backward direction quickly increases to a saturation value $j_0 \exp(-eV_c/kT)$.

The intermediate case of a natural barrier which is comparable in thickness with the mean free path has been investigated by Herzfeld[4] to show that the current will differ from that of the pure diode theory, Eq. (5), by less than five percent in the range ± 1 volt.

The forward characteristics predicted by all of these theories are much the same and the current may be approximately represented by

$$i = i_0[\exp(eV/kT)-1], \qquad (6)$$

where i_0 may vary slowly with voltage and is strongly dependent on temperature and contact potential.

The backward characteristics range from the saturation type of Eq. (5) to the approximately ohmic type of Eq. (3). These will be modified in the direction of increased currents by image force and tunneling effects.

A decision between the thick or thin natural barrier, and between exhausted or reserve region theories can be made from measurements of the Hall coefficient, R, and resistivity, ρ, as functions of temperature. From these data one may calculate the mean free path l; the depth, ΔE, of the donator levels below the bottom of the conduction band or of acceptor levels above the top of the full band in the case of N and P semi-

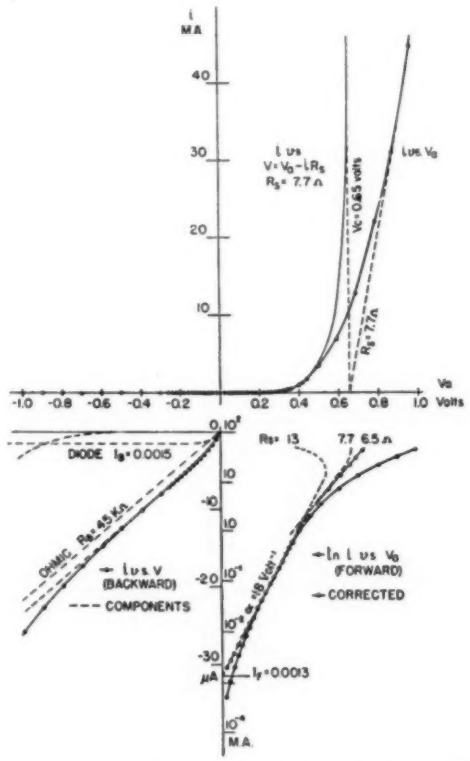

FIG. 2. Typical rectifier characteristic and analysis (W point on Ge, 300°K). Above, linear plot. Below, logarithmic plot of forward and enlarged linear plot of backward characteristic, showing method of correcting the former for spreading resistance and the analysis of the latter into its components.

[2] H. A. Bethe, NDRC Division 14 Report 43–12, MIT November 23, 1942.

[3] R. G. Sachs, NDRC Report 14–129, PB5190 Purdue, September 10, 1942 and PB5195 June 15, 1943.

[4] K. F. Herzfeld, NDRC Division 14 Report, PB5199 (May 5, 1944).

TABLE II. Electrical parameters of Ge and Si rectifiers.

Sample	Type	T°K	α volt^{-1}	$s=\alpha kT/e$	I_F ma	I_B ma	R_B 10^3 ohms
5Id-4	Ge-W	243	22	0.46	3×10^{-4}	4×10^{-4}	254
		295	20	0.52	3×10^{-3}	3.2×10^{-3}	77
		394	8.9–6.2	0.30–0.21	0.25–4.0	0.24	5.8
		549	7.0–5.0	0.33–0.24	2.0–3.0	0.3–0.5	0.09
5Id-2	Ge-W	286	7.4	0.18	0.20	0.25	1.1
		436	6.3	0.24	1.0	0.35	2.0
		506	4.9–2.8	0.21–0.12	2.0–4.0	0.60	0.2
Ge-35	Ge-Pt Ir	165	17.0	0.24	1.3×10^{-4}	$<5\times10^{-5}$	$>1.2\times10^{3}$
		239	16.6	0.34	7.0×10^{-4}	$<1\times10^{-4}$	$>1.6\times10^{3}$
		250	17.0	0.36	9.0×10^{-4}	1×10^{-4}	91
		269	18.0	0.41	1.1×10^{-3}	4.0×10^{-4}	78
		298	16.4	0.42	3.5×10^{-3}	1.6×10^{-3}	39
		299	17.5	0.45	3.0×10^{-3}	1.5×10^{-3}	46
		333	15.9	0.45	8.0×10^{-3}	6.0×10^{-3}	25
		351	17.0	0.51	8.0×10^{-3}	4.7×10^{-3}	31
		360	15.5	0.48	1.5×10^{-2}	1.2×10^{-2}	18
		393	15.3	0.52	3.0×10^{-2}	2.1×10^{-2}	13
GE-40	Ge-Pt Ir	167	19.2	0.28	2.0×10^{-5}	$<5\times10^{-5}$	>300
		241	16.2	0.33	5.0×10^{-4}	$<5\times10^{-5}$	>100
		251	17.7	0.38	3.0×10^{-4}	$<5\times10^{-5}$	>100
		271	18.2	0.42	4.0×10^{-4}	$<1\times10^{-4}$	167
		298	18.3	0.47	7.0×10^{-4}	$<1\times10^{-4}$	120
		299	19.1	0.49	6.0×10^{-4}	$<2\times10^{-4}$	82
		299	18.3	0.47	7.0×10^{-4}	$<1\times10^{-4}$	100
		333	17.7	0.50	2.0×10^{-3}	5.0×10^{-4}	74
		351	17.9	0.54	3.3×10^{-3}	1.1×10^{-3}	46
		360	17.3	0.53	4.4×10^{-3}	2.0×10^{-3}	53
		394	15.8	0.53	1.5×10^{-2}	7.5×10^{-3}	24
9 Ha	Ge-W	298	7.8	0.20	0.80	0.013	32
		358	11.9	0.36	0.71	0.28	5.0
		391	11.5	0.38	0.95	0.87	1.7
		426	9.0	0.33	2.5	3.0	0.50
		468	6.6	0.27	5.5	7.1	0.15
		521	7.0	0.31	10.2	7.8	1.9×10^{-3}
		574	4.4	0.22	22		7.0×10^{-2}
267	Ge-Pt Ir	298	14.4	0.37	6.0×10^{-3}	2.2×10^{-3}	16
172	Ge-Pt Ir	298	17.7	0.45	7.0×10^{-5}	$<1\times10^{-5}$	$>2\times10^{3}$
462	Ge-Pt Ir	298	34.7	0.88	7.0×10^{-5}	$<1\times10^{-5}$	$>2\times10^{3}$
2A-1′	Ge-W	298	10.6	0.27	2.4×10^{-2}	3.0×10^{-3}	7.2
2A-2′	Ge-W	298	14.6	0.37	2.0×10^{-2}	2.0×10^{-3}	>8.3
2A-3′	Ge-W	298	13.9	0.35	7.4×10^{-3}	$<2\times10^{-3}$	>11.9
2A-4′	Ge-W	298	11.7	0.30	5.0×10^{-2}	4.0×10^{-3}	3.2
2A-5′	Ge-W	298	10.7	0.27	3.0×10^{-2}	3.0×10^{-3}	4.6
2A-2	Ge-W	298	10.4	0.27	1.1×10^{-2}	2.5×10^{-3}	11.2
2A-3	Ge-W	298	12.0	0.31	3.0×10^{-2}	5.0×10^{-3}	3.0
2A-4	Ge-W	298	11.6	0.30	2.6×10^{-2}	5.0×10^{-3}	3.7
2A-5	Ge-W	298	13.3	0.34	9.0×10^{-3}	3.0×10^{-3}	6.8
2A-6	Ge-W	298	13.1	0.33	2.6×10^{-2}	5.0×10^{-3}	5.7
6Rb-1	Ge-W	298	22.0	0.56	7.0×10^{-3}	2.0×10^{-3}	14.3
6Rb-2	Ge-W	298	14.4	0.37	4.9×10^{-2}	1.4×10^{-2}	2.2
C-4	Si-W	298	18.4	0.47	6.0×10^{-3}	$<1\times10^{-3}$	9.4
C-6	Si-W	298	18.4	0.47	5.0×10^{-3}	1×10^{-3}	18.7
BTH-E1	Si-W	298	21.0	0.54	2.5×10^{-2}	$<5\times10^{-3}$	>3.2
BTH-E2	Si-W	293	16.4	0.41	1.0×10^{-2}	3.0×10^{-3}	7.0
BTH-E3	Si-W	177	11.0	0.17	2.0×10^{-2}	$<2\times10^{-3}$	3.9

conductors, respectively; the electron density, n; the density of donator or acceptor levels, N, and by Eq. (1) the natural barrier thickness, D. These data for several samples of silicon and germanium at 300°K are abstracted in Table I from the work of K. Lark-Horovitz and collaborators.[5] The thickness, D, is calculated for $V_c=0.3$ volt. The distance x_0 over which the potential barrier rises by kT/e is included, since this may be the more nearly correct critical distance to be compared with 1.[2]

In germanium the mean free path is three to ten times x_0, and in silicon approximately equal to x_0. In view of the work of Herzfeld,[4] the diode theory should, therefore, apply to these materials if the rectification is due to a natural barrier.

[5] K. Lark-Horovitz, "Preparation of semi-conductors and development of crystal rectifiers," NDRC Report 14–585, Purdue University.

That an insulator on the surface, or a depletion layer, is necessary for rectification in these materials seems highly improbable, since the general features of their characteristic curves are not at all sensitive to surface treatment. A freshly broken surface of a homogeneous single crystal may give characteristics quite the same as does a suitably polished and etched surface. There are surface treatments in use which could result in a depleted or insulating layer, but these treatments are not essential to the rectification. In any case, characteristics of the type of Eq. (6) are to be expected in at least the forward direction.

Experimental characteristics of typical silicon and germanium point contact rectifiers are shown in Fig. 2. In the forward direction the plot is made semi-logarithmic as well as linear to facilitate comparison with Eq. (6).

The original current data is plotted against the applied voltage V_a, and this voltage corrected to $V=V_a-iR_s$ where R_s is the spreading resistance in series with the contact. Since the effective diameter of the contact cannot be determined with adequate precision to calculate this correction satisfactorily, it is made by trial to give as regular a corrected curve as possible. According to the theories, as represented in Eq. (6), the corrected semi-log plot should be linear for $V \gg kT/e$ ($kT/e=1/40$ volt at 300°K). It is nearly always possible to find a value of R_s (6.5 ohms, Fig. 2) consistent with that calculated from the measured diameter, which will give a linear corrected curve in the usual range of V_a. As is indicated in the figure, however, this linear portion is usually not a continuation of the lower part where the spreading resistance is negligible. The upper slope for normal rectifiers tends to be somewhat less than the slope of the lower part. This is particularly true for germanium. If one attempts to make the upper slope as large as the lower by using a larger value of R_s, (13 ohms, Fig. 2) the corrected curve will usually begin to bend backward to give a negative differential resistance at applied voltages in the neighborhood of one volt. With some R_s between these values one could expect to obtain a corrected curve which approaches a vertical line corresponding to $V=V_c$ as an asymptote. Such a curve is approximated in Fig. 2 with $R_s=7.7$ ohms. This is equivalent to constructing the asymptote of differential resistance 7.7 ohms on the linear plot, cutting the voltage axis at V_c.

To make satisfactorily accurate determinations of R_s and V_c by either of these constructions would require data to higher voltages than is usually attainable because of erratic behavior of the contact. The interpretation of such data is also uncertain due to the strong dependence of the semi-conductor resistivity on temperature. In the case of germanium, the linear plots frequently show no definite region of truly constant slope, but the slope continues to increase as the voltage is raised to values for which it is certain there will be disturbances caused by heating.‡‡ This corresponds in the above analysis to a decrease in the effective R_s and an increase in effective V_c as the applied voltage is increased. In view of these difficulties, perhaps the best one can say at present is that the straight line correction on the semi-log plot gives a minimum value of R_s, and that which produces a negative differential resistance gives a maximum value appropriate to the voltage range used. These uncertainties at the upper part of the characteristic do not affect it appreciably below 0.4–0.5 volts, however, so that comparison with theory can be made in this range.

According to Eq. (6), if a current i_0 is added to the observed forward current, a straight line of slope e/kT should be obtained on a semi-logarithmic plot. Usually it is possible, as is indicated in Fig. 2, to find such a current which will yield a linear region of considerable length. Let us denote this current which corresponds to i_0, by I_F in order to distinguish it from a similar current, I_B, obtained from the backward characteristic. Also denote the slope of the corrected curve by α.

The curves of Fig. 2 are quite typical of silicon and germanium rectifiers. While linear regions of ln$i$ extending over five powers of ten in current have been found, a range of one hundred to one thousand is more common. Characteristics are found which deviate from this linear behavior, usually in the sense of two oppositely curved portions, convex upward at the low currents and convex downward at the higher currents, but these are the exception.

These characteristics are in qualitative agreement with the prediction of theory, but the quantitative agreement is very poor. The result common to all theories discussed above is that the logarithmic slope should be very approximately $\alpha=e/kT=40$ volts^{-1} at 300°K. The experimental slopes are always less than this. The value shown in Fig. 2 of approximately one half e/kT is typical as is evident from the data listed in Table II. Occasional higher values of α do occur, but the one of .88 e/kT quoted for sample 462 is exceedingly rare. Much less uncommon are values down to one-fifth or one-sixth e/kT. Use of Pt-Ir alloy for the metal in contact with germanium has some tendency to increase the value of α as compared with that obtained with tungsten.

It is evident that the approach to the linear portion of the curve is much less rapid than is indicated by the theories, so that the data can usually be represented by a characteristic of the form $i=I_F[\exp(\alpha V)-1]$.

The wide range of I_F indicated in the table is not inconsistent with the diode theory. One may set $I_F=i_0A=Aj_0\exp(-eV_c/kT)$ where A is the area of the contact. Using 10^{-6} cm^2 as an average value of A and with $j_0 \simeq 10^6$ amp/cm^2 computed from the Hall

‡‡ This apparent inconstancy of spreading resistance is under investigation, particularly for the "high back voltage" type rectifier. See reference 5; Ralph Bray, Phys. Rev. **74**, 1218 (1948).

coefficient measurements, one finds a range of I_F from 1 ma to 0.2μa corresponds to a variation of V_c between 0.17 and 0.39 volt. These are reasonable limits, not in disagreement with the estimates to be made from the spreading resistance correction discussed above.

Granting, however, the I_F is correct, it follows from the exponential dependence of the current on α, that with α one-half e/kT the current at 0.3 volts is $\frac{1}{4}$ percent of the theoretical value, and for α one-third e/kT, it is only 1/30 percent of theoretical. Clearly some modification of the theories is necessary.

A similar disagreement has long been known in the low voltage region; e.g., Wagner, Schottky and Spenke[1]. Although for a reserve semi-conductor one has flexibility in α (Schottky α_1) by the factor of $\frac{1}{3}$, values for Cu_2O, MoS, TeO_2, etc., are found as low as $\frac{1}{3}e/kT$. It may be possible in the low voltage region to account for the remaining discrepancies by the assumption of essentially ohmic imperfections in the barrier. In the higher voltage region, this solution is not possible, since the minimum resistance which can be assigned from the back characteristic will contribute nearly insignificant currents to the forward characteristic at voltages above about 0.1 volt. In the rather favorable example of Fig. 2, if currents equivalent to the entire back current are subtracted from the forward characteristic, the value of α is increased from 18 to only 20.4 volts^{-1}.

The back direction characteristic in Fig. 2 has been plotted on an enlarged linear scale. It is evident that the current does not saturate as required by the simple diode theory, but in general it may be analyzed into a diode component, an ohmic component, and a rapidly increasing component or "tail." The latter part is frequently nearly exponential in character and may, therefore, be ascribed to tunneling through a thin barrier, as suggested by the reversed rectification in this case, Eq. (2), or it may be due to the effect of image force in lowering the barrier. These processes will be discussed below.

The diode-like component, I_B, and the resistance of the ohmic component, R_B, are listed in Table II. It is evident, contrary to the diode theory, that I_B is not the same as I_F. It may be larger or smaller, but is usually smaller, particularly at low temperature. Determination of a new I_F from the forward characteristic after correcting for the ohmic contribution to it, as discussed above, does not in general bring the two currents into coincidence.

The existence of the important ohmic component of the back current suggests that the barrier is of the thick depleted layer type leading to Eq. (3). From the data given and the discussion above, however, it seems highly unlikely that this is necessarily the case. It would rather seem that irregularities which have an essentially ohmic behavior must exist in the thin natural barrier.

The diode theory, with the assumption of such ohmic regions, combined with some form of tunneling and image force barrier lowering should give agreement with the back direction characteristic. Introduction of such modifications also affects the forward characteristic. The ohmic regions themselves will not appreciably reduce the discrepancy between the theoretical and experimental logarithmic slope, as was shown above, but the other effects might be expected to do so.

The effect of barrier lowering by image force has been investigated by Bethe[2] and others[3, 6] and the tunnel effect by Courant.[6] The effect of image force is to lower the barrier by an increasing amount as the voltage is reduced in the forward direction or increased in the backward direction. The tunnel effect may be expressed as an additional lowering.

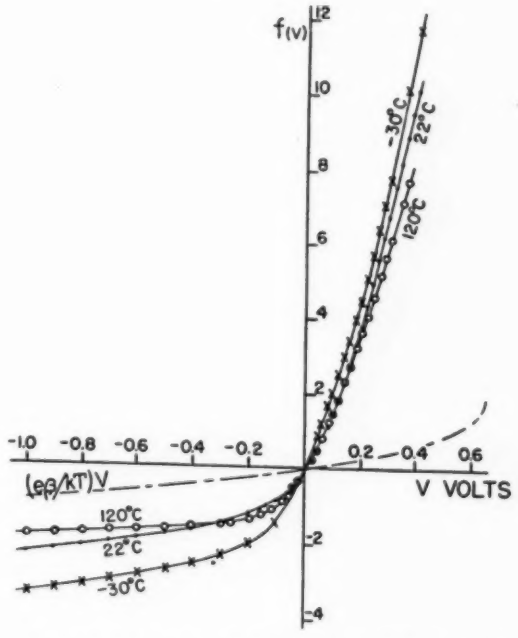

FIG. 3. Typical behavior of the exponent $f(V)$ in a corrective term, $\exp[-f(V)]$, to bring the simple diode theory into agreement with experiment (Data of 5Id-4, Table II). Dashed curve shows the type of correction afforded by the barrier lowering modification of the theory.

The total potential lowering from the original value V_c, may be expressed as $\Delta V_c=\beta(V_c-V)$ where β[Courant $(1-\theta)$] is positive and nearly constant for $V\ll V_c$, and increases slowly as V approaches V_c. Subtracting this correction from the V_c of Eq. (5) gives

$$j=j_0\exp[-e(1-\beta)V_c/kT]\exp(-e\beta V/kT)\times[\exp(eV/kT)-1] \quad (7)$$

instead of Eq. (5). For comparison with experiment,

[6] E. D. Courant, NDRC Division 14 report MIT May 17, 1943. See also Phys. Rev. **69**, 684 (1946). In this abstract the sign of the corrective exponent should be positive. The impression which might be given by the abstract that there is agreement between the corrected theory and experiment, is misleading. The full paper shows that the essential discrepancy still exists, as discussed herein.

instead of Eq. (6) one should then use

$$i = i_0' \exp(e\beta V/kT)[\exp(eV/kT) - 1], \qquad (8)$$

where i_0' is nearly constant except for $V \simeq V_c$.

As was mentioned earlier, this additional factor is of the correct form to agree with the observed "tail" in the back direction. For negative V the variation of the term in brackets is negligible and the back current increases exponentially. As nearly as one can divorce this effect from the diode and ohmic parts, the experimental data require values of β of approximately 0.01 to 0.05. It seems possible to obtain values of this order from the theory.

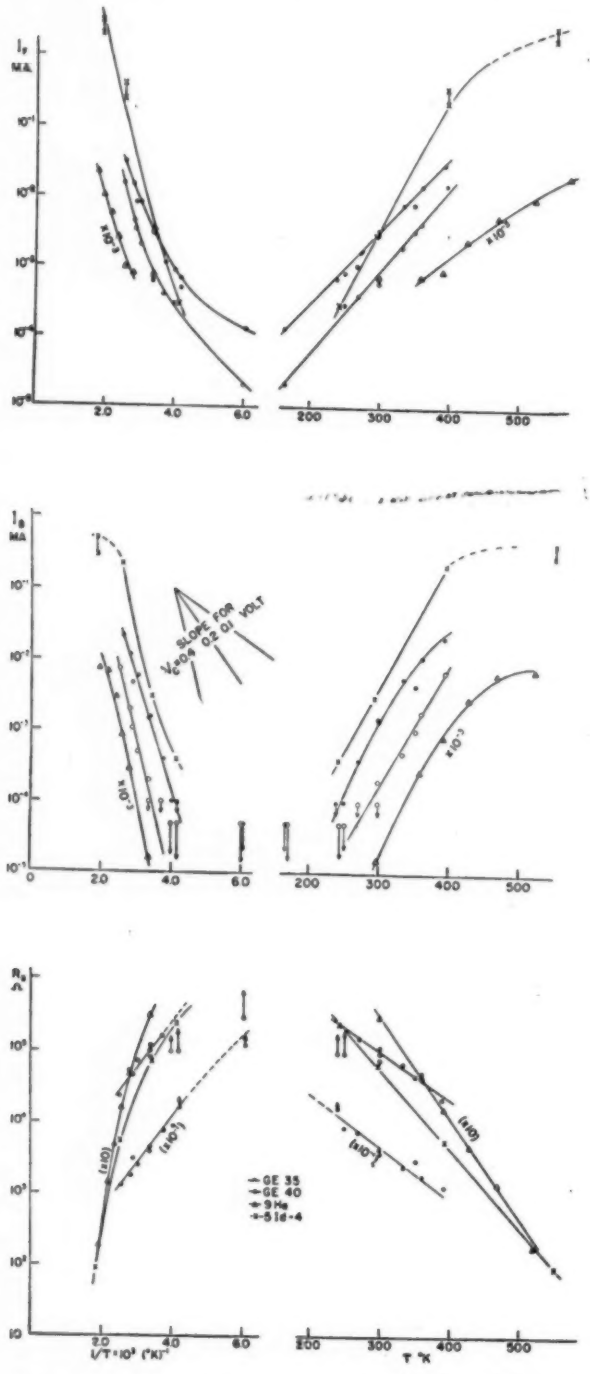

FIG. 4. Plots of I_F, I_B, and R_B against $1/T°K^{-1}$ and $T°K$; variation with reciprocal of temperature is not always linear and slopes of I_F and I_B are not equal.

In the forward direction there would also seem to be an improvement over the simple theory. On expanding Eq. (8)

$$i = i_0'\{\exp[e(1-\beta)V/kT] - \exp[-e\beta V/kT]\} \qquad (9)$$

and since for $V \gg e/kT$ the second term is negligible, one obtains an exponential characteristic whose logarithmic slope has been reduced from e/kT to $(1-\beta)e/kT$.

The reduction possible by this correction cannot, however, be made sufficient to agree with experiment. The values of α usually found range downward from approximately $0.5e/kT$. This requires values of β of 0.5 and larger. For $\beta = 0.5$ the predicted value of $\alpha = (1-\beta)e/kT = 0.5e/kT$, but in this case Eq. (9) shows that the backward characteristic, V negative, is identical with the forward characteristic and there would be no true rectification. The required values of β larger than 0.5 would give a reversal of sign of rectification. Moreover, quantitative estimates¶ of the magnitude of β show that it cannot be as large as 0.5, except for very low values of V_c.

To be in agreement with experiment, in the terminology of Eq. (8), β would have to increase from ≈ 0.1–0.05 for negative V, to ≈ 0.5–0.7 or more for positive V. While the theories give a variation in this sense, it is negligible in amount compared with the required variation.

The experimental data may be used to determine an empirical function, $-f(V)$, to replace $-e\beta V/kT$ in the corrective term of Eq. (8). The function can be determined only to within an additive constant, because of the uncertainties in the area of contact and contact potential included in i_0'. The curves of Fig. 3, for a germanium rectifier at three different temperatures, have, therefore, been arbitrarily displaced to pass through the origin. In contrast with this rapidly varying function required by experiment, the simple diode theory gives a constant value for $f(V)$ and the barrier lowering theories give a curve which slowly increases in slope as the voltage is increased toward V_c.

The characteristics of a number of crystals have been obtained at various temperatures. Considerable difficulty is encountered in maintaining the same contact over a wide range of temperatures. Some of the more reliable data are listed in Table II.

At the high temperatures, analysis according to Fig. 2 is difficult, because the spreading resistance uncertainty affects a large part of the forward characteristics and also because the measurements in the back direction do not always extend far enough to differentiate between ohmic current and diode current, due to the

¶ References 5 and 6; also V. F. Weisskopf, NDRC Report 14–133 (May 12, 1943).

extremely rapid increase of the latter with temperature. At the low temperatures the back current may become too small for accurate determinations. In these cases ranges of values are given.

The general behavior is that α decreases and I_F and I_B increase very rapidly with increasing temperature, as would be expected from the theories. R_B decreases with increasing temperature, so that whatever mechanism is assumed to provide this ohmic component, it must have an effective resistivity which is very temperature sensitive. The large changes with temperature are demonstrated in the $f(V)$ plots of Fig. 3, remembering that $f(V)$ is the argument of an exponential.

Examination of the values of α in Table II shows that these rectifiers are of different types. For 5Id2, 5Id4, and 9Ha, the ratio of α to the simple theory value, $s=\alpha kT/e$, in general falls with increase of temperature. There is some indication of an initial rise with a maximum at $\approx$400°K. GE-35 and 40, on the other hand, show a nearly constant value of α, so that the ratio s rises with temperature. Physically, GE-35 and 40 differ from the others in that a condenser discharge had been passed through the contact and the junction was thought to be at least partially welded.||

The variations of I_B, I_F, and R_B with temperature are shown in semi-logarithmic plots as functions of T and $1/T$ in Fig. 4. The apparent disturbance at the highest temperatures may be due partially to physical changes in the contact, although the data were reasonably reproducible. In spite of the limited nature of the data, there seems to be a somewhat better agreement with an equation of the form $\exp(CT)$ than with $\exp(-K/T)$, particularly for I_F, whereas I_F and I_B should each be equal to i_0, and yield a negative exponential in $1/T$, according to Eq. (5). The slope of the log plot against $1/T$ should be $-eV_c/k$ except for possible variation in n, and the small effect of the $T^{\frac{1}{2}}$ coefficient. (The Hall effect data indicate that n is nearly constant over the range of these measurements.) The order of magnitude of V_c required by the data is from 0.1 to 0.65 volts. This range obtained from the temperature variation is reasonably consistent with the range of 0.17 to 0.39 volts estimated earlier from I_F itself. There is, however, no detailed correlation. The values of A and n involved in i_0 are certainly of the same order of magnitude for these samples and a large value of i_0 (I_F or I_B), corresponding to a small V_c should be associated with a small negative slope on the $1/T$ diagram. The lines should tend to radiate from a common point (Aj_0) on the $1/T=0$ axis.

From this discussion it becomes evident that a satisfactory theory of the silicon or germanium rectifier must allow the logarithmic slope of the forward characteristic to be reduced below e/kT by a factor which is in itself temperature dependent. The intercepts corresponding to I_F and I_B must not be required to be equal, and must vary approximately exponentially with T. The theory must also provide a mechanism for an ohmic current, as well as a diode component, in the back direction, and the resistance of this component must vary exponentially with temperature, instead of having the relatively slow variation of the semiconductor resistivity.

Part II will be concerned with the development of a diode theory to meet these conditions. The development is based upon a proposal of H. A. Bethe that one abandon consideration of the contact as being a perfectly homogeneous one with all parts having the same contact potential, and adopt the point of view that it consists of more or less isolated regions having lower contact potentials than their surroundings. These regions should be considered to be very small spots ranging in contact potential from very low values up to (or beyond) the values normally expected.

It is essential to consider that the current through any such spot is limited by the local spreading resistance in series with it, or by the fact that the current density through it has a maximum value of j_0. A graphical construction will quickly show that with suitable distributions of the number and size of spots having a given contact potential, one can obtain an α of any value less than e/kT. The temperature dependence of α may be attained by varying the distribution functions. An ohmic component is provided by the contributions of the spots having zero or negative contact potential, for which there is no barrier, and its temperature dependence by the variation with temperature of the number or size of these spots.

|| These experimental crystals were kindly supplied to us by Dr. H. Q. North, General Electric Company, Schenectady, New York.

Electric Breakdown in CO_2 from Low Pressures to the Liquid State*

D. R. YOUNG†
Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received September 26, 1949)

The electric breakdown in CO_2 has been investigated from low pressures through the critical point into the liquid state. Paschen's similarity law is verified for low pressures. At high pressures small departures are observed for long gap-lengths and large departures for small gap-lengths. Simultaneously the scatter of the breakdown voltage becomes independent of illumination and the breakdown strength dependent on the cathode material; this appears due to the onset of field emission. Measurements of prebreakdown currents have yielded values for Townsend's first coefficient as well as for the field emission constants. For small gap-lengths the prebreakdown currents are higher than the normal field emission equation predicts, indicating some new process effective at short gap-lengths. The transition from the gaseous to the liquid state does not produce a discontinuous change of the breakdown voltage.

INTRODUCTION

THREE experimental variables describe the d.c. breakdown characteristics of gases in uniform fields after the gas type, the electrode material, and the source of primary ionization have been chosen: the breakdown voltage v, the gas density ρ, and the electrode spacing d. Paschen established experimentally,[1] that the breakdown voltage of a spark gap is a function of the product ρd only, $v=f(\rho d)$. This equation, in which Paschen used pressure instead of the density, is called the Paschen similarity law. It states that the voltage drop per free path is the important variable.

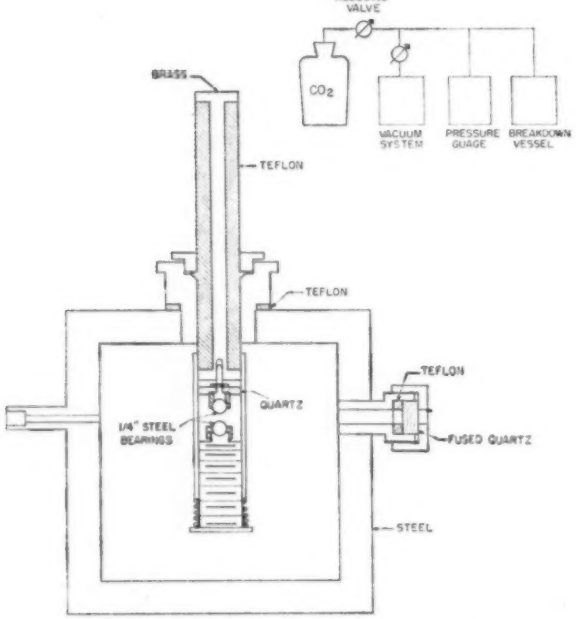

FIG. 1. Schematic diagram of breakdown equipment.

This law cannot hold at high pressures as von Hippel has pointed out.[2] The breakdown field strength in atmospheric air amounts to about 3×10^4 v/cm. For solids and liquids, with the density several thousand times higher, the breakdown strength has risen only to the order of 10^6 v/cm. Hence, instead of about 1.5 v per free path in gases, only 0.01–0.1 v is required for breakdown over the corresponding path lengths in condensed phases. It seems that the breakdown mechanism has changed from the Townsend type, in which the ionization is held in check by electronic excitation processes, to a new mechanism in which the vibration states of the solids and liquids appear to play a decisive role. This latter mechanism has been the object of investigation in this laboratory for some time.[3,4] It is the aim of the research presented here to learn more about the transition region between the breakdown types in gases and in liquids.

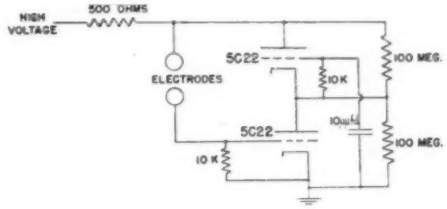

FIG. 2. Electrode protection circuit.

Carbon dioxide was used as the test material since its critical point is readily accessible and its equation of state well known.

EQUIPMENT AND MEASURING TECHNIQUES FOR BREAKDOWN IN THE GASEOUS STATE

A schematic diagram of the apparatus used for the breakdown measurements is shown in Fig. 1. The equipment was evacuated before the CO_2 was admitted. For measurement of pressure a mercury column was used in the lower pressure range (up to two atmospheres)

* This work was sponsored jointly by the ONR, the Army Signal Corps, and the Air Force under ONR Contract N5ori-07801.

† From a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Physics at the Massachusetts Institute of Technology.

[1] F. Paschen, Wied. Ann. **37**, 69 (1889).

[2] A. von Hippel, Zeits. f. Physik **80**, 19 (1933).

[3] A. von Hippel, J. App. Phys. **8**, 815 (1937).

[4] A. von Hippel and R. S. Alger, Phys. Rev. **76**, 127 (1949).

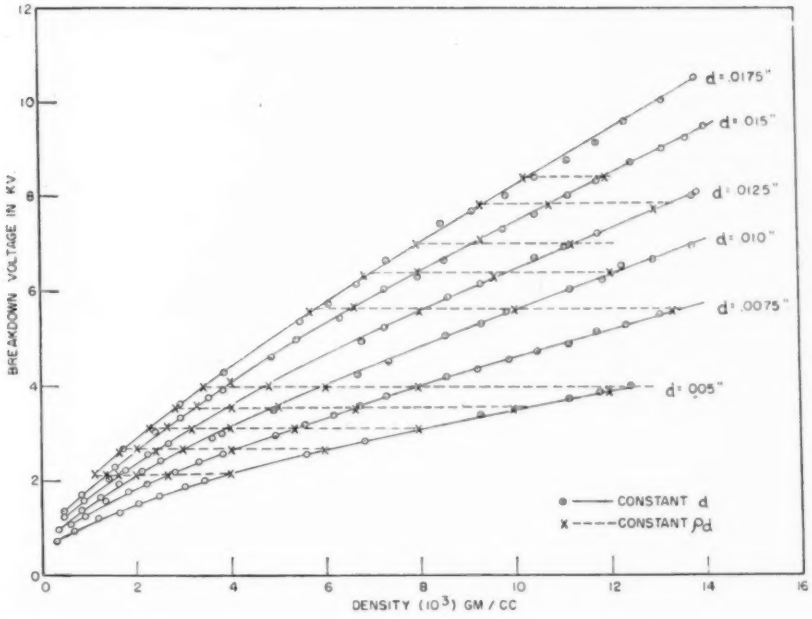

FIG. 3. Breakdown voltage *vs.* density with gap-length as parameter.

and Clapp pressure gauges, calibrated by a dead-weight tester, above this limit.

Initially the high voltage insulator of the breakdown vessel was made of polystyrene, but it proved soluble in liquid CO_2 and was therefore replaced by Teflon. Steel ball bearings 0.250 in. diameter served as electrodes since their polish is excellent and their low cost makes them expendable.

Since preliminary experiments indicated that the insulating materials used were not rigid enough to insure an accurate separation if they had to support the pressure, the electrode holder was designed as shown in Fig. 1. The spacing was adjusted by the spring-loaded screw on the lower electrode, and calibration marks 0.0005 in. apart were included. It was possible to adjust the spacing with an accuracy of about ± 0.0002 in.

Teflon was used for the pressure seals with excellent results if the gaskets were entirely confined. A quartz window allowed illumination of the cathode by ultraviolet light.

Without current limitation each breakdown damages the electrode surface and requires its reconditioning. To avoid this damage to the electrodes, the high voltage was removed about 10^{-7} sec. after initiation of the breakdown, by using two hydrogen thyratrons (Sylvania 5C22) in series, triggered by the breakdown current (Fig. 2). Two thyratrons were required since the voltages used exceeded the rating for a single tube.

The high voltage was measured with a multi-range current amplifier used in conjunction with a 10^9-ohm standard resistance. The standard resistance was made up of 100 10^7-ohm one percent resistors which were mounted in 15 copper shells stacked in series. Measure-

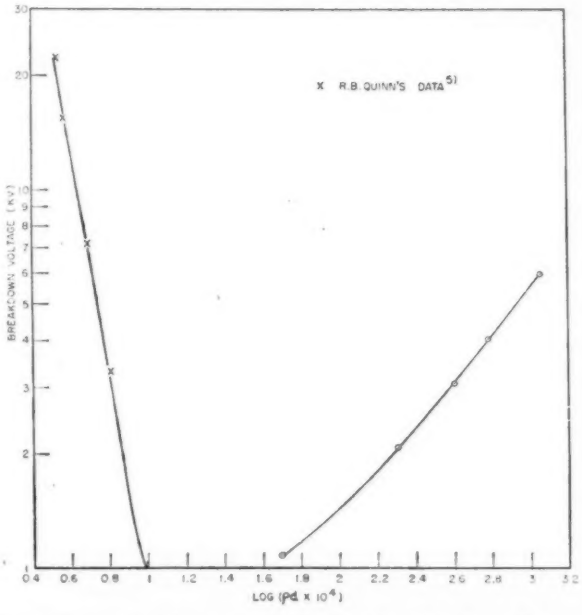

FIG. 4. Breakdown voltage *vs.* ρd (ρ in g/cc, d in mils).

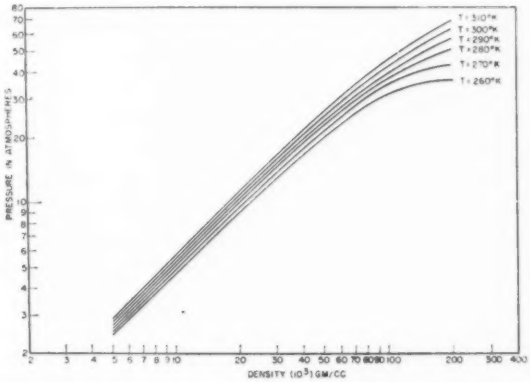

FIG. 5. Pressure-density relationship from Beattie-Bridgman equation.

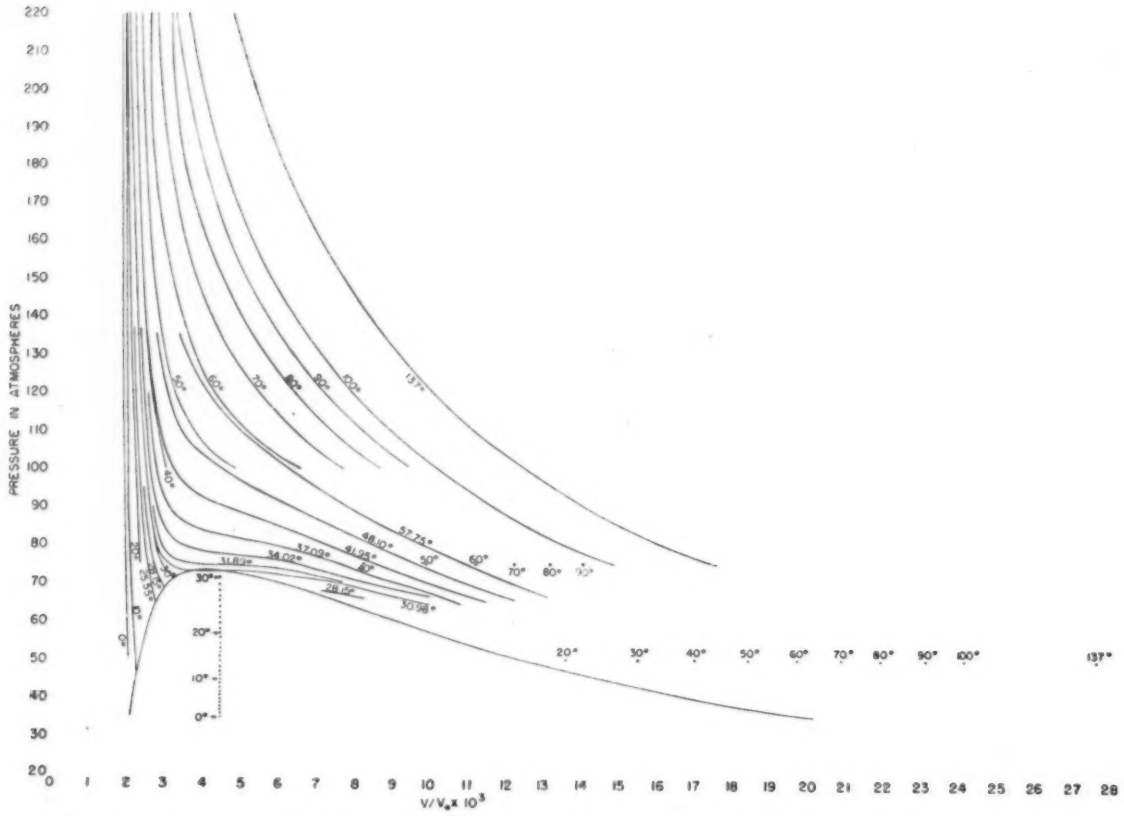

FIG. 6. Pressure-volume data from *International Critical Tables* (see reference 7) in vicinity of critical point.

ments of the over-all resistance indicated that it was accurate to within one percent. The shell mounting reduced the field surrounding the resistors and thus prevented corona from this cause.

To reduce the scatter in breakdown voltage due to a scarcity of initiating electrons, the cathode was illuminated by ultraviolet light from a mercury arc. The ratio of electrode diameter to gap-length was kept large enough to assure uniform field geometry.

BREAKDOWN MEASUREMENTS AT LOW GAS PRESSURES

In Fig. 3 the measured breakdown voltage is plotted as a function of gas density with the gap-spacing as parameter. In this region the ideal gas law holds: $P=1.865T\rho$ where P is given in atmospheres, ρ in g/cm, and T in degrees K.

It can be seen that the curves of constant $d\rho$ are horizontal lines, indicating that Paschen's law is obeyed in this region. Consequently, the data of Fig. 3 can be combined into a single characteristic of breakdown voltage *vs.* ρd. Figure 4 shows this curve and its extension to lower pressures taken from the paper of Quinn.[5]

[5] R. B. Quinn, Phys. Rev. **55**, 482 (1939).

BREAKDOWN MEASUREMENTS AT HIGH PRESSURES

For this pressure range the most accurate equation of state is the Beattie-Bridgman equation[6]

$$P=\frac{RT(1-\epsilon)}{V^2}(V+B)-\frac{A}{V^2}, \qquad (1)$$

where $A=A_0(1-a/V)$, $B=B_0(1-b/V)$, and $\epsilon=c/VT^3$,

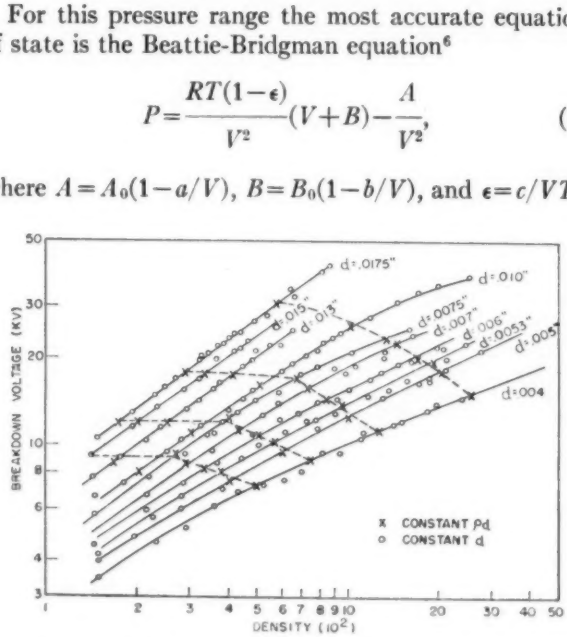

FIG. 7. Breakdown voltage *vs.* density with gap-length as parameter.

[6] J. A. Beattie and O. C. Bridgman, J. Am. Chem. Soc. **49**, 1665 (1927).

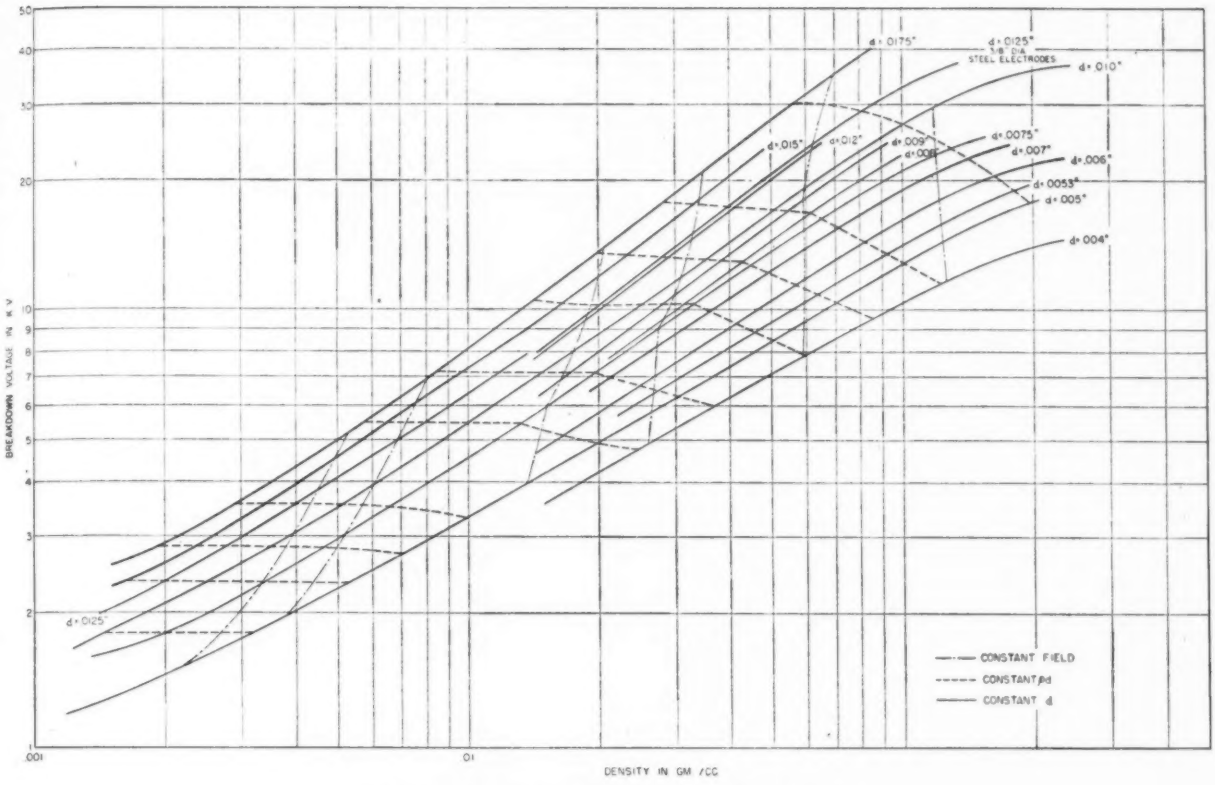

FIG. 8. Breakdown *vs.* density for various gap-lengths.

with P given in atmospheres, V in cc/g, T in degrees K. For CO_2, $R=1.865$, $B_0=2.381$, $b=1.6443$, $A_0=2586.0$, $a=1.6210$, and $C=15\times10^6$. According to Eq. (1) the ideal gas law is valid at room temperature to within one percent for pressures up to three atmospheres. Figure 5 shows the pressure-density relation according to the Beattie-Bridgman equation for various temperatures. In the vicinity of the critical point the relation between pressure and density must be established experimentally. Pertinent data are given in the *International Critical Tables*[7] as plotted in Fig. 6; the relative volume V/V_0 (abscissa) where V_0 refers to the volume at $T=0°C$, and $P=$one atmosphere. The density is correspondingly

$$\rho=\frac{1}{(508V/V_0)}\ [\text{g/cc}]. \qquad (2)$$

The range covered by these measurements extends from densities of 0.01 to 0.24 g/cc (that is, to the density of the vapor above the liquid at room temperature) (Fig. 7).

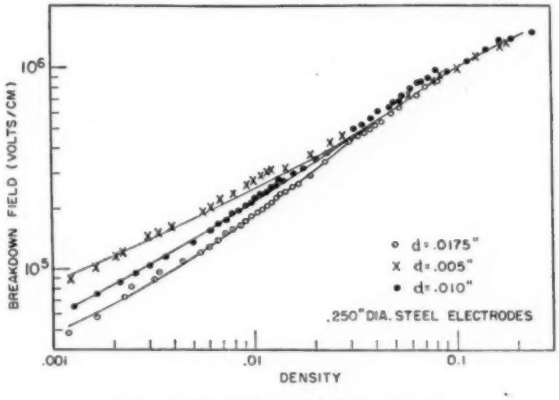

FIG. 9. Breakdown field *vs.* density.

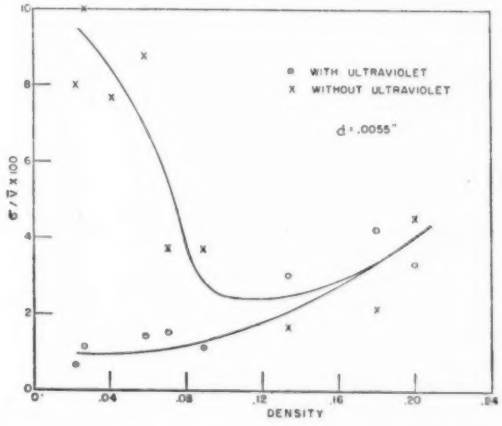

FIG. 10. Effect of illumination on scatter of breakdown voltage measurements.

[7] *International Critical Tables* (McGraw-Hill Book Company, Inc., New York, 1926).

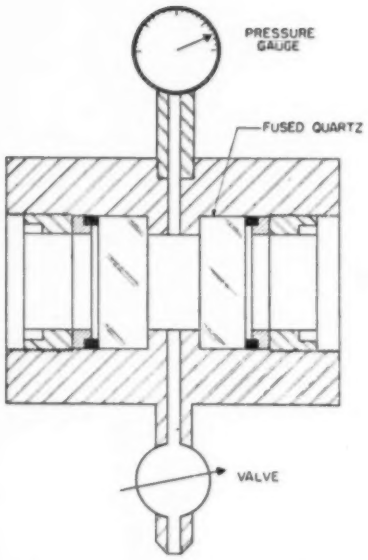

FIG. 11. High pressure absorption cell.

For long gap-lengths small departures from Paschen's law are observed as the deviation from the horizontal of the $d\rho$=constant curves indicates; for small gap-lengths these departures are much larger.

In Fig. 8 the high and low pressure data are combined to show the systematic increase of the deviations. In addition to the ρd curves, lines of constant field strength are indicated. It is seen that for small gap-lengths and high densities these lines are vertical, indicating that the breakdown voltage is determined by the average field strength for a given density, independent of the gap-spacing. If the breakdown field strength is plotted as a function of density, the lines for different gap-spacings consequently converge for high densities (Fig. 9).

If the standard deviation σ is plotted as a function of the density, with the cathode either illuminated or not by ultraviolet light, the curves shown in Fig. 10 are obtained. It is evident that, for the higher densities, the ultraviolet light has no effect.

The ineffectiveness of illumination might be caused by an increase of the ultraviolet absorption with gas density which would prevent the light from reaching the cathode. Since no absorption data were available for CO_2 covering this pressure range, measurements were made using a Cary recording spectrophotometer and the gas chamber shown in Fig. 11. The quartz windows were selected for transparency and had a cut-off limit near 2000A. There was no measurable absorption up to the saturation pressure at room temperature for wave-lengths down to 2000A, which rules out this explanation.

Another possible explanation for the data in Fig. 10 is that field emission provides the necessary initiating electrons. Breakdown measurements were therefore made using zinc-coated cathodes because zinc has a relatively low work function. They were prepared by evaporating zinc on steel ball bearings previously heated to 350°C in a high vacuum in order to obtain the same quality of surface as in the previous experiments. In Fig. 12 the results are compared with those using steel cathodes. At high gas densities the breakdown voltage is appreciably reduced by using zinc-coated electrodes, as would be expected.

At the higher densities the breakdown voltages for

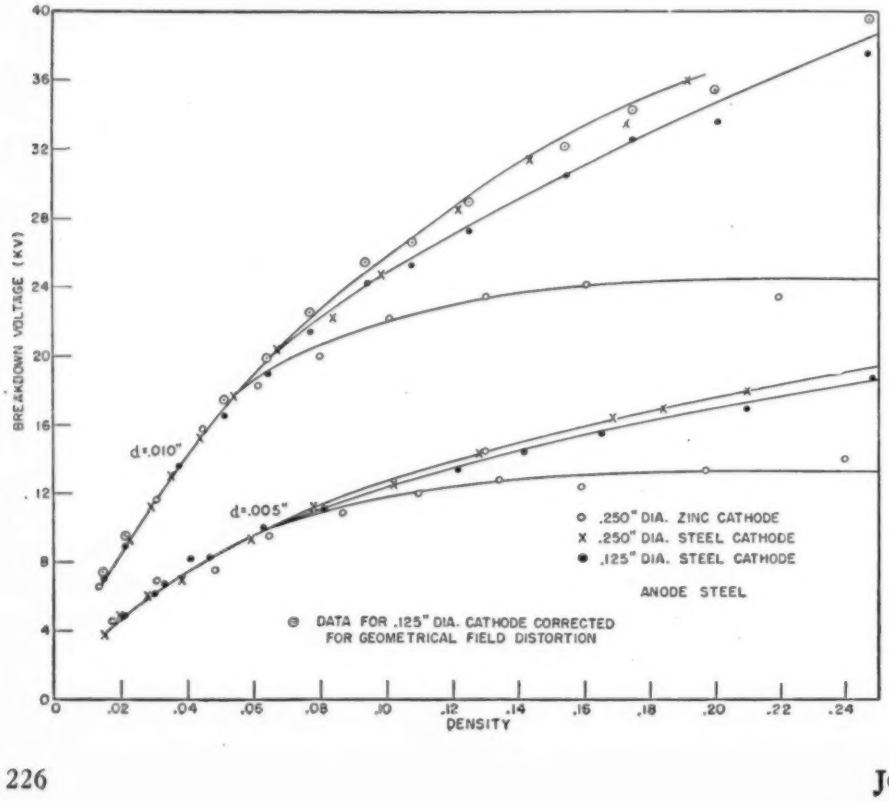

FIG. 12. Breakdown voltage measurements comparing various cathodes.

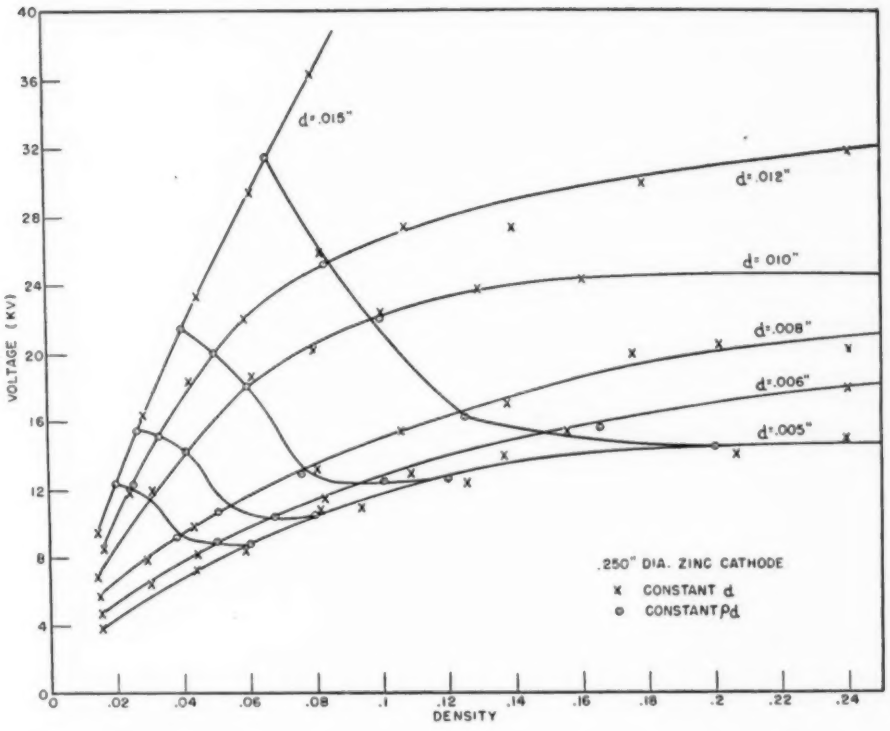

FIG. 13. Breakdown voltage *vs.* density for zinc cathode with gap-length as parameter.

0.125-in. diameter steel electrodes lie lower than those for the 0.250-in. steel electrodes. This is due to the geometrical field distortion for these electrode configurations. The data for the 0.125-in. balls, corrected for geometrical field distortion, yield the same results as for 0.250-in. cathodes (cf. Fig. 12). The breakdown voltage is not influenced by changing the anode metal or by using a 0.125-in. diameter anode instead of the 0.250-in. one.

In Fig. 13 additional breakdown curves for zinc cathodes are given which illustrate the very early failure of Paschen's law.

CURRENT MEASUREMENTS AT HIGH PRESSURES

The prebreakdown currents that flow, when only the α- and γ-type processes of Townsend are present, are proportional to the cathode illumination and are not self-sustaining. If the current is initiated by field emission, there will be a resulting current that does not depend on cathode illumination. A study of prebreakdown currents was initiated, therefore, to check the "field emission hypothesis" suggested above.

A single-stage current amplifier using a Victoreen Instrument Company VX-41A-type tube was constructed to measure the current. The amplifier was operated entirely by self-enclosed batteries with associated meters and adjustable resistors so that changes in battery voltage could be compensated for without change in the calibration of the amplifier (Fig. 14). The indicating instrument used was a 0–10 microammeter. The zero drift was negligible and the calibration remained constant over long periods of time.

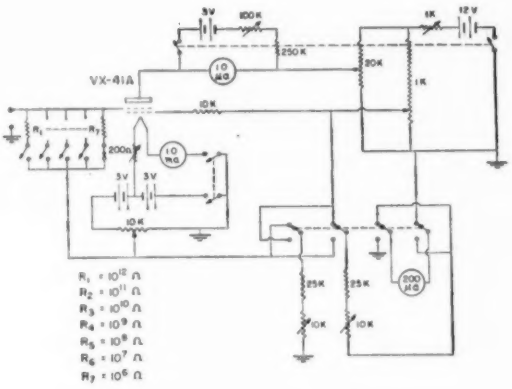

FIG. 14. Current amplifier.

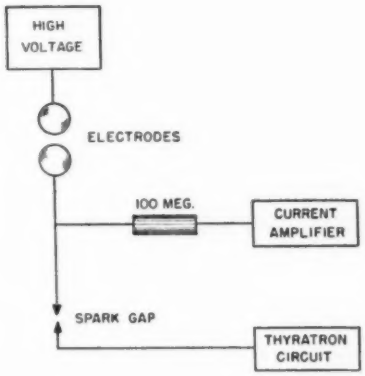

FIG. 15. Protection circuit used in conjunction with current amplifier.

The battery drain was small enough to keep the rate of change of the battery voltage negligible for several hours.

In order to use the protecting circuit described earlier, it was necessary to provide for a triggering voltage in a manner not disturbing to the current measurements (Fig. 15). If breakdown should occur, the small gap fires and supplies the necessary triggering voltage for current limitation and protection of the instrument.

The displacement current charging the electrode capacity of the breakdown gap caused considerable trouble. Instabilities in the high voltage supply, causing displacement current to flow, made the current readings erratic even though the line voltage was regulated by a Sorensen constant voltage transformer. To minimize this difficulty the measurements were made at night. During the course of a particular run (at each

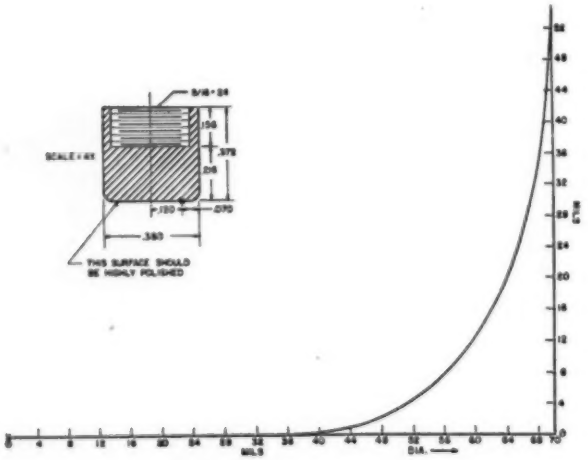

FIG. 16. Electrodes shaped according to Rogowsky's curves (see reference 8).

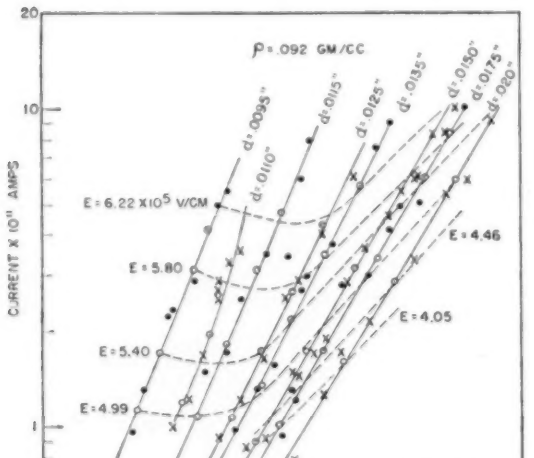

FIG. 17. Prebreakdown current *vs.* applied voltage for density of 0.092 g/cc.

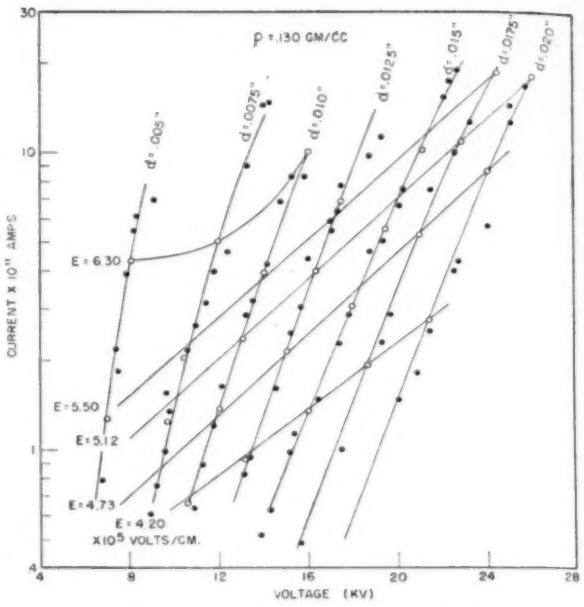

FIG. 18. Prebreakdown current *vs.* applied voltage for density of 0.130 g/cc.

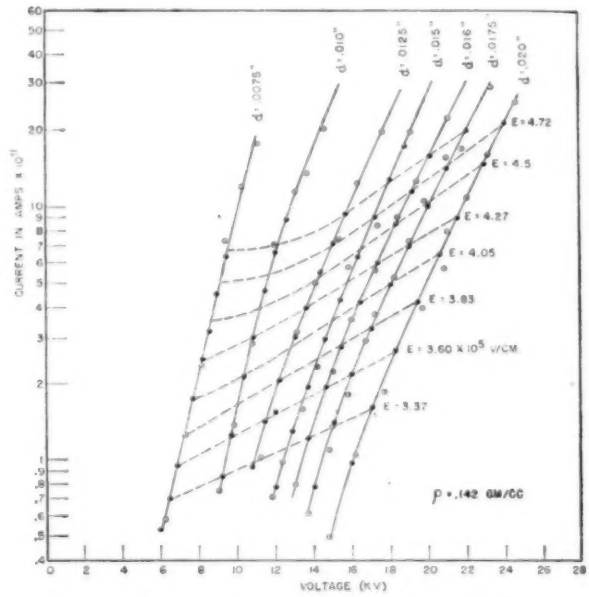

FIG. 19. Prebreakdown current *vs.* applied voltage for density of 0.142 g/cc.

setting of the applied voltage) it was necessary to wait until the electrode capacity charged up to the new voltage before a reading could be taken. These waiting times amounted to several minutes in the case of the smallest currents. A more sentitive indicating instrument would allow the use of smaller grid resistances and reduce this trouble.

The breakdown vessel was directly attached to the amplifier to eliminate disturbances introduced by connecting cables, and provision was made to bring out an insulated connecting wire from the lower electrode

through a pressure seal of the same construction as used at the high voltage terminal (cf. Fig. 1).

Steel electrodes shaped to give a uniform field‡ were used to provide a more definite surface area than that offered by spherical electrodes (Fig. 16).[8] Designed for a maximum spacing of 0.020 in., the electrodes then serve also for smaller spacings.

As other workers have already reported,[9] field emis-

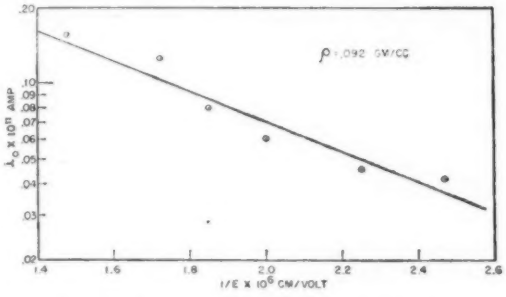

FIG. 20. i_0 *vs.* $1/E$ for density of 0.092 g/cc.

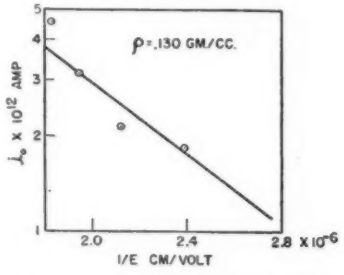

FIG. 21. i_0 *vs.* $1/E$ for density of 0.130 g/cc.

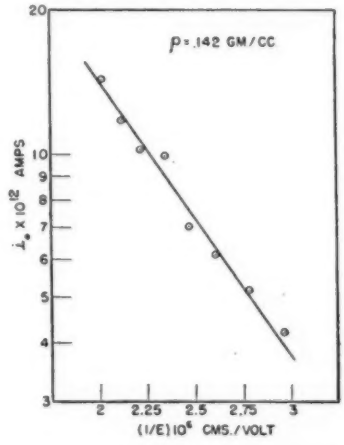

FIG. 22. i_0 *vs.* $1/E$ for density of 0.142 g/cc.

sion currents are very sensitive to the nature of the surface and the electrodes have to be conditioned. The main conditioning procedure consisted in sparking, with the protection circuit preventing excessive burning of the electrodes.

‡ Made and polished by the A. D. Jones Optical Company, Cambridge, Massachusetts.

[8] W. Rogowski, Archiv. f. Elektrotechnik **12**, 1 (1923).

[9] R. A. Millikan and C. F. Eyring, Phys. Rev. **27**, 51 (1926).

TABLE I. Measured values for the field emission coefficients and for α as a function of the applied field.

	ρ=0.092 g/cc E	α
	4.05×10^5 v/cm	71.5 cm^{-1}
a=1.21×10^{-11} amp.	4.46	83.9
	4.99	91
b=1.35×10^6 v/cm	5.4	97
	5.8	98.5
	6.22	106.2
	ρ=0.130 g/cc E	α
a=4.2×10^{-11} amp.	4.2×10^5 v/cm	54 cm^{-1}
	4.73	74
b=1.32×10^6 v/cm	5.12	77
	5.50	85
	ρ=0.142 g/cc E	α
	3.37×10^5 v/cm	25.8 cm^{-1}
	3.60	32.2
a=2.1×10^{-10} amp.	3.83	37.5
	4.05	44.1
	4.27	44.3
b=1.345×10^6 v/cm	4.5	52
	4.72	56.4
	4.95	58.75

Combining the usual equation for field emission with the Townsend equation for impact ionization one obtains:

$$i = ae^{-b/E}e^{\alpha d}. \tag{3}$$

In Figs. 17 to 19 the logarithm of the current is plotted as a function of the voltage for different gap-lengths and curves of constant field strength are drawn. It is noted that straight lines result in this logarithmic plot as anticipated, but only for the longer gap-lengths. It is possible to calculate the Townsend first coefficient

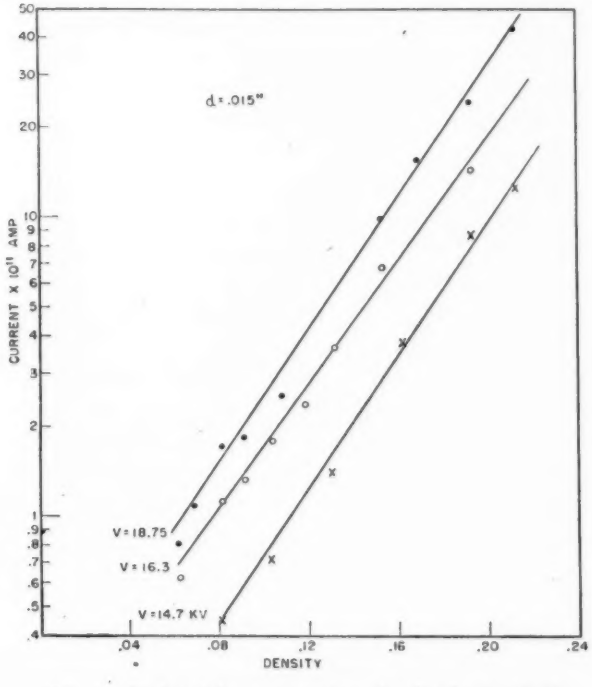

FIG. 23. Prebreakdown current *vs.* density for constant voltage and gap-length.

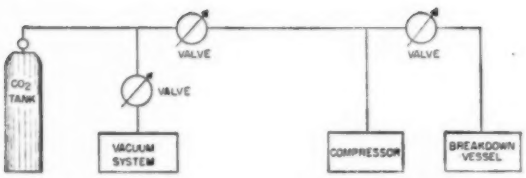

FIG. 24. Experimental apparatus for liquid breakdown.

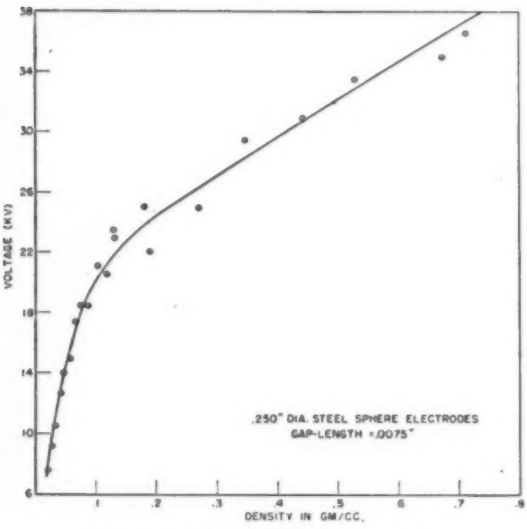

FIG. 25. Breakdown voltage *vs.* density.

α for different strengths from these lines. Then by extrapolation of the lines of constant E to zero voltage (which is the same as zero d) one obtains

$$i_0 = ae^{-b/E} = i_0(E) \quad \text{(cf. Eq. 3)}, \tag{4}$$

and a plot of i_0 as a function of $1/E$ gives the constants a and b (Figs. 20–22). Again a straight line results showing that also the exponential dependence on $1/E$ is fulfilled. Thus the current measurements demonstrate that field emission occurs under these high pressure conditions.

The values of a, b, and α obtained for these gas densities are given in Table I.

If the applied field strength is constant, the current increases with increasing density as shown by Fig. 23. Since α decreases, and b is constant, this is due to an increasing a, as already seen in Table I.

For small gap-spacings the current is larger than predicted by Eq. (3) as can be seen in Figs. 17 to 19, since the curves of constant field strength are no longer straight lines.

BREAKDOWN IN LIQUID CO_2

To make measurements in the liquid phase, a special compressor was constructed to compress the CO_2 gas which is delivered from the tank at saturation pressure. Alternatively, the breakdown vessel was cooled to about 1°C below the tank temperature, in order to condense the CO_2. This latter method is not as flexible as the first one mentioned for establishing a desired pressure above the liquid phase. A diagram of the experimental apparatus is shown in Fig. 24. To study the effect of pressure on the breakdown strength, a thermostat-controlled oil bath was constructed in which the breakdown vessel could be immersed. (No measurements were made in the liquid-vapor equilibrium region where two phases are present.) Measurements were made at the saturation pressure, heating from room temperature and with either gas or liquid CO_2 filling the breakdown vessel while the vessel was sealed off to prevent density change. In both cases there was no change in breakdown voltage with temperature. Therefore, if the density is constant for a given gap-length the breakdown voltage is independent of temperature in the range covered (for gas, 25°–55°C; for liquid, 25°–35°C).

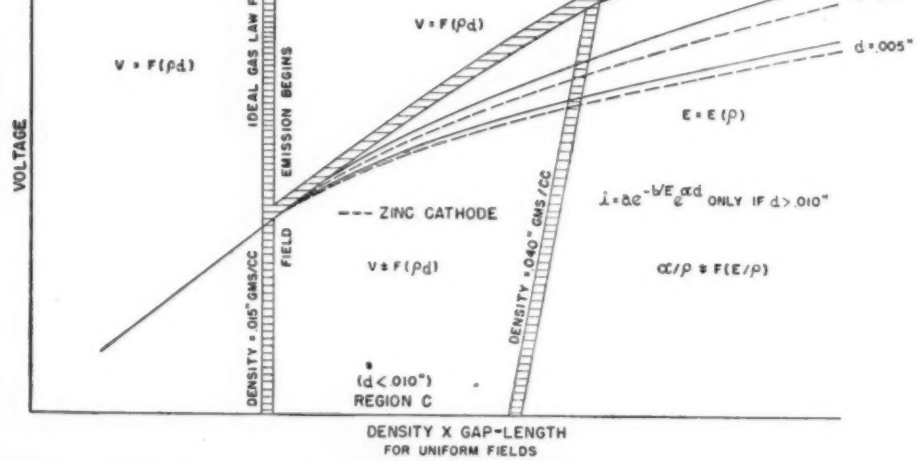

FIG. 26. Breakdown voltage *vs.* ρd (schematic).

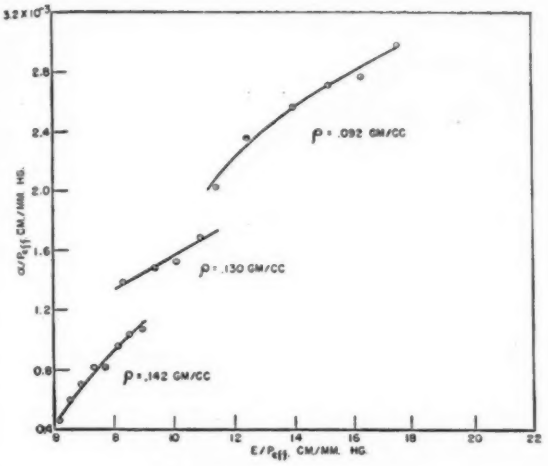

FIG. 27. α/P_{eff} *vs.* E/P_{eff}.

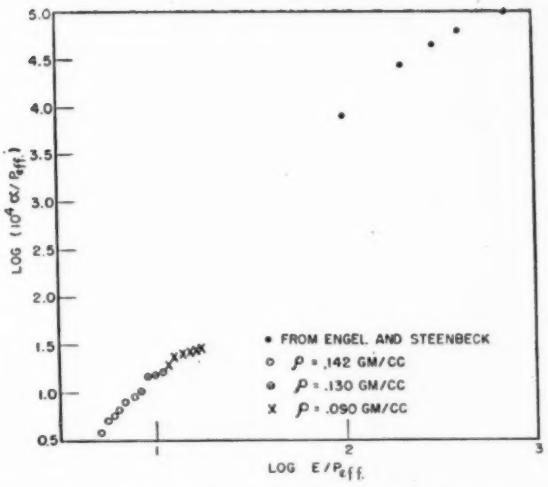

FIG. 28. Present data compared with that given by Engel and Steenbeck (see reference 10) for $\log_{10}(10^4\alpha/P_{eff})$ *vs.* $\log_{10}E/P_{eff}$.

By heating the liquid to 33°C, that is, 2° above the critical temperature, it was possible to change the density continuously from that corresponding to the liquid to that of the vapor. A continuous variation in the breakdown voltage results (Fig. 25).

DISCUSSION

Breakdown Measurements

The most interesting result of these breakdown measurements is the discontinuity in the slope of the ρd curves (Fig. 8) which becomes pronounced above densities $\simeq$0.015 g/cc. This effect corresponds to the branching of the breakdown voltage characteristics plotted in Fig. 26 which occurs at $\rho=0.015$ g/cc. The large departures from Paschen's law for smaller gap-lengths indicate the importance of a new process that aids the breakdown but does not satisfy the similarity law requirements. The onset of this process is partially determined by the cathode material (cf. Fig. 8 with Fig. 13), and its presence is also revealed by the pre-breakdown current which was larger than predicted by Eq. (3) for d's $<$0.010 in. (Figs. 17 to 19).

The explanation of this effect is not obvious. One possibility is that space charge becomes important for small gap-lengths. This would increase the effective ionization coefficient and, in addition, raise the field at the cathode which increases the probability of secondary emission of electrons (γ-process) and the field emission. A solution of the field equation does not predict that an appreciable field distortion occurs for current densities averaged from the measured current; however, field emission does not occur uniformly, and therefore the current densities may be very much higher than assumed.

Current Measurements

At low pressures the relation $\alpha/\rho=f(E/\rho)$ holds, but at high pressures, as Fig. 27 indicates, α/ρ is not uniquely determined by E/ρ. This might be expected because the interaction of the molecules will increase the energy loss suffered by the electrons through the excitation of vibrations.

In order to make comparisons with low pressure data, it is convenient to define an effective pressure as

$$P_{eff}=P_0\rho/\rho_0=3.87\times10^5\rho \quad [\text{mm Hg}] \qquad (5)$$

if P_0 and ρ_0 are measured at S.T.P., where the ideal gas law is valid. Low pressure data for α/P *vs.* E/P given by Engel and Steenbeck[10] are combined with the present data in Fig. 28.

The probability of impact ionization before breakdown decreases rapidly with increasing gas density at the highest density studied (Fig. 27).

The increase of the field emission constant a with increasing density may be due to the increasing dielectric constant and its effect on the image force field of the escaping electrons.

The author is grateful to Professor A. von Hippel for suggesting this problem and for constant advice and encouragement during the progress of the work. The many helpful suggestions given by members of the High Voltage Group of this laboratory are also appreciated.

[10] M. Steenbeck and A. v. Engel, *Elektrische Gasentladungen* (Verlag Julius Springer, Berlin, 1932), Vol. 1, p. 106.

A Method for the Numerical Calculation of Hydrodynamic Shocks

J. VonNeumann and R. D. Richtmyer
Institute for Advanced Study, Princeton, New Jersey
(Received September 26, 1949)

The equations of hydrodynamics are modified by the inclusion of additional terms which greatly simplify the procedures needed for stepwise numerical solution of the equations in problems involving shocks. The quantitative influence of these terms can be made as small as one wishes by choice of a sufficiently fine mesh for the numerical integrations. A set of difference equations suitable for the numerical work is given, and the condition that must be satisfied to insure their stabilty is derived.

I. INTRODUCTION

IN the investigation of phenomena arising in the flow of a compressible fluid, it is frequently desirable to solve the equations of fluid motion by stepwise numerical procedures, but the work is usually severely complicated by the presence of shocks. The shocks manifest themselves mathematically as surfaces on which density, fluid velocity, temperature, entropy and the like have discontinuities; and clearly the partial differential equations governing the motion require boundary conditions connecting the values of these quantities on the two sides of each such surface. The necessary boundary conditions are, of course, supplied by the Rankine-Hugoniot equations, but their application is complicated because the shock surfaces are in motion relative to the network of points in space-time used for the numerical work, and the differential equations and boundary conditions are non-linear. Furthermore, the motion of the surfaces is not known in advance but is governed by the differential equations and boundary conditions themselves. In consequence, the treatment of shocks requires lengthy computations (usually by trial and error) at each step, in time, of the calculation.

We describe here a method for automatic treatment of shocks which avoids the necessity for application of any such boundary conditions. The approximations in it can be rendered as accurate as one wishes, by suitable choice of interval sizes and other parameters occurring in the method. It treats all shocks, correctly and automatically, whenever and wherever they may arise.

The method utilizes the well-known effect on shocks of dissipative mechanisms, such as viscosity and heat conduction.[1] When viscosity is taken into account, for example, the shocks are seen to be smeared out, so that the mathematical surfaces of discontinuity are replaced by thin layers in which pressure, density, temperature, etc. vary rapidly but continuously. Our idea is to introduce (artificial) dissipative terms into the equations so as to give the shocks a thickness comparable to (but preferably somewhat larger than) the spacing of the points of the network. Then the differential equations (more accurately, the corresponding difference equations) may be used for the entire calculation, just as though there were no shocks at all. In the numerical results obtained, the shocks are immediately evident as near-discontinuities that move through the fluid with very nearly the correct speed and across which pressure, temperature, etc. have very nearly the correct jumps.

It will be seen that for the assumed form of dissipation (and, indeed, for many others as well), the Rankine-Hugoniot equations are satisfied, provided the thickness of the shock layers is small in comparison with other physically relevant dimensions of the system. We then consider one way in which the transition from differential to finite-difference equations can be made and we discuss the mathematical stability of these equations. It will be seen that the dissipative terms have the effect of making the stability condition more stringent than the familiar one of Courant, Friedrichs, and Lewy,[2] but not seriously so if the amount of dissipation introduced is only enough to produce a shock thickness comparable with the spatial interval length of the network used.

The method has been applied, so far, only to one-dimensional flows, but appears to be equally suited to the study of more complicated flows; where, indeed, shock calculations by direct application of the Hugoniot equations would ordinarily be prohibitively difficult, even for rapid, automatic computers.

The discussions which follow are primarily intended to give a complete picture of the ideas and mathematical procedures involved. In some places (Chapter VII, also the essential inferences from some of the material of Chapters IV, V) the mathematical discussions are, however, carried through only with a view to give a complete chain of the necessary procedure, but not with all the detail that rigorous proofs in a primarily mathematical paper would require. The reason for doing this was partly desire to avoid inconvenient length, partly that of not wishing to have to select now the precise degree of generality for the

[1] Lord Rayleigh (Proc. Roy. Soc. **A84**, 247 (1910)) and G. I. Taylor (Proc. Roy. Soc. **A84**, 371 (1910)) showed, on the basis of general thermodynamical considerations, that dissipation is necessarily present in shock waves. Later, R. Becker (Zeits. f. Physik **8**, 321 (1922)) gave a detailed discussion of the effects of heat conduction and viscosity. Recently, L. H. Thomas (J. Chem. Phys. **12**, 449 (1944)) has investigated these effects further in terms of the kinetic theory of gases.

[2] Courant, Friedrichs, and Lewy, Math. Ann. **100**, 32 (1928). It is in this important paper that these authors first published their discovery of the conditional stability of the difference-equation integration method for partial differential equations.

validity of the method. It seems preferable to reserve such discussions for later occasions.

The validity of our methods has been tested empirically on various computational applications. These are partly still under analysis, and will be published and discussed elsewhere.

II. THE BASIC EQUATIONS

Consider a one-dimensional fluid motion. Let x be the Lagrangean coordinate, and $X=X(x, t)$ be the Eulerian coordinate. That is, $X(x, t)$ gives the position, at time t, of a fluid element that was initially at position x. Let $\rho_0(x)$ be the initial density, so that V and U, given by

$$V(x, t)=(1/\rho_0(x))(\partial X/\partial x) \tag{1}$$

and

$$U(x, t)=\partial X/\partial t, \tag{2}$$

are the specific volume and fluid velocity, respectively. The equations of motion, of energy, and of continuity are:

$$\rho_0(\partial U/\partial t)=-(\partial/\partial x)(p+q), \tag{3}$$

$$(\partial \mathcal{E}/\partial t)+(p+q)(\partial V/\partial t)=0, \tag{4}$$

and

$$\rho_0(\partial V/\partial t)=(\partial U/\partial x). \tag{5}$$

In these equations, $p=p(x, t)$ is the ordinary (or static) fluid pressure and $\mathcal{E}=\mathcal{E}(x, t)$ is the internal energy per unit mass. A connection between $\mathcal{E}$, p, V is established by an equation of state, which will be assumed, for the purpose of illustration, to have the form

$$\mathcal{E}=(pV)/(\gamma-1) \tag{6}$$

which holds, for example, in the case of a perfect gas. γ is a constant >1. It is supposed that the dissipative mechanism can be represented by the additional term q in the pressure, which is assumed to be negligibly small, except in the neighborhood of the shock.

III. THE EXPRESSION FOR q

The dissipation is introduced for purely mathematical reasons. Therefore q may be taken as any convenient function of p, V, etc. and their derivatives, provided that the following requirements are met:

1. The Eqs. (3), (4), and (5) must possess solutions without discontinuities.
2. The thickness of the shock layers must be everywhere of the same order as the interval length Δx used in the numerical computation, independently of the strength of the shock and of the condition of the material into which it is running.
3. The effect of the terms containing q in (3) and (4) must be negligible outside of the shock layers.
4. The Hugoniot equations must hold when all other dimensions characterizing the flow are large compared to the shock thickness.

We shall show that the expression

$$q=-\frac{(\rho_0 c\Delta x)^2}{V}\frac{\partial V}{\partial t}\cdot\left|\frac{\partial V}{\partial t}\right| \tag{7}$$

meets the requirements. c is a dimensionless constant near unity. The dissipative mechanism is essentially a non-linear viscosity as can be seen more clearly if (7) is written equivalently (for the one-dimensional case) as

$$q=-\frac{(c\Delta x)^2}{V}\frac{\partial U}{\partial x}\cdot\left|\frac{\partial U}{\partial x}\right| \tag{8}$$

by use of (5). To show that the expression (7) meets the stated requirements, we consider a steady-state shock.

IV. STEADY-STATE PLANE SHOCK

Imagine a long pipe containing a fluid initially in equilibrium (thermally and mechanically), into which a piston is pushing from one end with constant speed, as shown in Fig. 1. In the absence of dissipation the specific volume, V, and the fluid velocity, U, are, at a given instant, as shown by the solid curves in the graphs, whereas in the presence of dissipation they are as shown by the dashed curves. In either case, the shock is steady, at least approximately, after it has gone to a sufficiently great distance from the initiating piston. Then U, V, $\mathcal{E}$, etc. depend on x and t only through the combination

$$w=x-st, \tag{9}$$

where s is the speed of the shock relative to the original, or Lagrangean, coordinates. We suppose that ρ_0 and Δx are constants (independent of x).

It is convenient to define

$$M=\rho_0 s \tag{10}$$

—in a co-moving coordinate system, M is the mass crossing unit area in unit time—whereupon Eqs. (3)–(5) become:

$$M(dU/dw)=(d/dw)(p+q), \tag{11}$$

$$(d\mathcal{E}/dw)+(p+q)(dV/dw)=0, \tag{12}$$

and

$$-M(dV/dw)=dU/dw. \tag{13}$$

Then, (11) and (13) give:

$$-M^2(dV/dw)=(d/dw)(p+q); \tag{14}$$

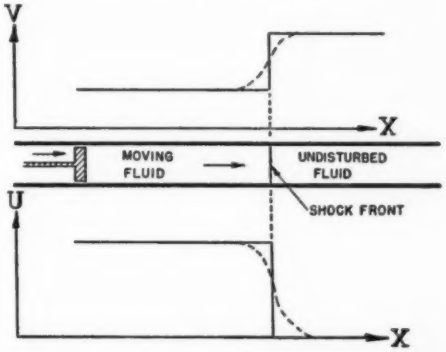

FIG. 1. Steady-state plane shock.

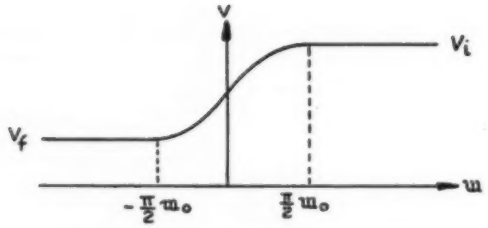

FIG. 2. Specific volume in steady shock.

and (12) and (14) give

$$\frac{d\mathcal{E}}{dw}+\frac{d}{dw}[(p+q)V]+M^2V\frac{dV}{dw}=0. \tag{15}$$

The solutions of (13), (14), (15) are:

$$MV+U=C_1 \tag{16}$$

$$M^2V+p+q=C_2 \tag{17}$$

$$\mathcal{E}+(p+q)V+\tfrac{1}{2}M^2V^2=C_3, \tag{18}$$

where C_1, C_2, and C_3 are constants.

Let the initial and final values be denoted by:

$$\text{As } w\to\infty\,;\quad V\to V_i,\quad p\to p_i,\quad \mathcal{E}\to\mathcal{E}_i,\quad q\to 0; \tag{19}$$

$$\text{As } w\to-\infty\,;\quad V\to V_f,\quad p\to p_f,\quad \mathcal{E}\to\mathcal{E}_f,\quad q\to 0. \tag{20}$$

Then (17) gives:

$$M^2(V_i-V_f)=P_f-P_i; \tag{21}$$

and (18) and (21) together give:

$$\mathcal{E}_f-\mathcal{E}_i=\tfrac{1}{2}(P_i+P_f)(V_i-V_f). \tag{22}$$

(21) and (22) are the equations of Hugoniot and are seen to be independent of the amount and form of the dissipation, provided that $q\to 0$ as $w\to\pm\infty$. The physical reason for this is that the Hugoniot equations are direct consequences of the conservation laws of mass, momentum and energy, and the form of dissipation assumed is such as top reserve the over-all conservation of these quantities. These laws require that in a shock a certain amount of mechanical energy be converted irreversibly into heat. In the steady state, the motion adjusts itself, in the shock layer, until precisely that amount of work is done against the pressure q according to Eq. (3) and is converted into heat according to Eq. (4).

To investigate the shape of the shock, we first look for solutions satisfying

$$(\partial V/\partial t)\leq 0 \quad\text{or}\quad (dV/dw)\geq 0. \tag{23}$$

This is normally the case for a shock moving to the right. Then (7) can be written:

$$qV=+(Mc\Delta x)^2(dV/dw)^2. \tag{24}$$

From (17) and (18)

$$\mathcal{E}-\tfrac{1}{2}M^2V^2=C_3-C_2V, \tag{25}$$

so that by (6),

$$pV=[(\gamma-1)/2]M^2V^2+C_4V+C_5, \tag{26}$$

whereupon (17) gives:

$$qV=C_2-[(\gamma+1)/2]M^2V^2-C_4V-C_5. \tag{27}$$

The right member of (27) is a quadratic in V that vanishes for $V=V_i$ and for $V=V_f$, so that, clearly,

$$(Mc\Delta x)^2\left(\frac{dV}{dw}\right)^2=qV=\frac{\gamma+1}{2}M^2(V_i-V)(V-V_f). \tag{28}$$

(Equation (28) can also be obtained from (27) by direct use of the Hugoniot equations and the equations that fix C_2, C_4, C_5.)

To solve (28), put:

$$\psi=V-\frac{V_i+V_f}{2},\quad \psi_0=\frac{V_i-V_f}{2},\quad \varphi=\frac{\psi}{\psi_0}, \tag{29}$$

so that

$$c\Delta x\frac{d\psi}{dw}=[(\gamma+1)/2]^{\frac{1}{2}}(\psi_0^2-\psi^2)^{\frac{1}{2}}, \tag{30}$$

or

$$c\Delta x\frac{d\varphi}{dw}=[(\gamma+1)/2]^{\frac{1}{2}}(1-\varphi^2)^{\frac{1}{2}}. \tag{31}$$

Therefore,

$$w=[2/(\gamma+1)]^{\frac{1}{2}}c\Delta x\int\frac{d\varphi}{(1-\varphi^2)^{\frac{1}{2}}}=w_0\arcsin\varphi \tag{32}$$

where

$$w_0=[2/(\gamma+1)]^{\frac{1}{2}}c\Delta x. \tag{33}$$

Finally,

$$\psi=V-\frac{V_i+V_f}{2}=\frac{V_i-V_f}{2}\sin\frac{w}{w_0}. \tag{34}$$

Because of our initial assumption (23) that $dV/dw\geq 0$, we can use only a half wave of the solution (34), but this half wave can be pieced together with two other particular solutions,

$$V\equiv V_i \quad\text{and}\quad V\equiv V_f \tag{35}$$

(they satisfy Eq. (28)), to make the composite continuous solution depicted in Fig. 2. w_0 is a measure of the thickness of the shock, and is of order Δx, provided c is of order unity, independently of the strength of the shock and conditions ahead of it, by Eq. (33). Throughout most of the system q is negligible in comparison with the ordinary pressure, p, because of the factor $(\Delta x)^2$ in (7); but in the shock layer q becomes comparable with p because of the abnormally large value of $\partial V/\partial t$ there.

Expression (7) thus meets all requirements.

It may be noted that if we had looked instead for solutions having $dV/dt>0$, no solution would have been found for which all quantities are continuous and bounded, because in that case the opposite sign would occur in (28), leading to the hyperbolic instead of ordi-

nary sine function. Therefore negative shocks do not exist in the steady state, for our equations. Negative shocks do not exist in the physical world, either, so that our expression (7) is satisfactory from this point of view.

V. STABILITY OF THE DIFFERENTIAL EQUATIONS

Suppose that at some instant there is superposed on a desired solution $U(x, t)$, $V(x, t)$, etc., a small perturbation δU, δV, etc. It is of interest to know whether the perturbation grows with increasing time. To investigate this we replace U, V, etc. by $U+\delta U$, $V+\delta V$, etc. in (3), (4), (5), (8), after first rewriting (4), by use of (6), in the form:

$$[\gamma p+(\gamma-1)q](\partial V/\partial t)+V(\partial p/\partial t)=0. \tag{36}$$

We obtain in this way the equations of first variation:

$$\rho_0(\partial/\partial t)\delta U=-(\partial/\partial x)(\delta p+\delta q); \tag{37}$$

$$\left.\begin{aligned}\frac{\partial V}{\partial t}[\gamma\delta p+(\gamma-1)\delta q]+[\gamma p+(\gamma-1)q]\frac{\partial}{\partial t}\delta V\\ +V\frac{\partial}{\partial t}\delta p+\frac{\partial p}{\partial t}\delta V=0,\end{aligned}\right\} \tag{38}$$

$$\delta q=\frac{(c\Delta x)^2}{V^2}\frac{\partial U}{\partial x}\cdot\left|\frac{\partial U}{\partial x}\right|\delta V-2\frac{(c\Delta x)^2}{V}\cdot\left|\frac{\partial U}{\partial x}\right|\frac{\partial}{\partial x}\delta U, \tag{39}$$

$$\rho_0\frac{\partial}{\partial t}\delta V=\frac{\partial}{\partial x}\delta U. \tag{40}$$

In writing the last term of (39) we have assumed that the perturbation is not large enough to alter the sign of $\partial U/\partial x$. (37) to (40) are a set of simultaneous linear differential equations for δU, δV, δp, δq. Their coefficients depend on the desired solution U, V, p, q, and are thought of as smoothly varying functions of x, t. We shall be concerned with rapidly varying perturbations. We therefore treat the coefficients of (37) to (40) as constants in a small region and look for solutions having the form:

$$\delta U=\delta U_0 e^{ikx+\alpha t},\quad \delta V=\delta V_0 e^{ikx+\alpha t},\text{ etc.,} \tag{41}$$

where δU_0, δV_0, δp_0, δq_0, k, and α are constant and k is real. Substitution of (41) into (37) to (40) leads to four simultaneous homogeneous linear equations in δU_0, δV_0, δp_0, δq_0. The vanishing of the determinant of these equations establishes an equation connecting α and k. By solving this equation, for given k, and examining the real part of α, we can determine whether a given Fourier component (41) of the perturbation grows with increasing time. This program is readily carried out, but we omit the details in the interest of brevity. The determinantal equation is:

$$\begin{aligned}(\alpha\rho_0)^2\gamma\frac{\partial V}{\partial t}+2\frac{\rho_0\alpha}{V}\frac{\partial V}{\partial t}(kc\Delta x)^2\left|\frac{\partial U}{\partial x}\right|\\ -\frac{1}{V^2}\frac{\partial V}{\partial t}(kc\Delta x)^2\frac{\partial U}{\partial x}\cdot\left|\frac{\partial U}{\partial x}\right|+k^2\alpha[\gamma p+(\gamma-1)q]\\ +\rho_0^2\alpha^3V+2\rho_0\alpha^2(kc\Delta x)^2\left|\frac{\partial U}{\partial x}\right|\\ -\frac{\alpha}{V}(kc\Delta x)^2\frac{\partial U}{\partial x}\cdot\left|\frac{\partial U}{\partial x}\right|+k^2\frac{\partial p}{\partial t}=0.\end{aligned} \tag{42}$$

If we restrict our attention to Fourier components with very large k, only certain terms of (42) need be retained, the others being negligible either by viture of being of lower order in k or by virtue of being of lower order in α. Two cases are distinguished: in shock regions we retain all terms in (42); in normal regions we drop the dissipative terms (those containing Δx). In the two regions, the dominant terms, in the sense explained, give: shock regions:

$$\alpha=-2\frac{(kc\Delta x)^2}{\rho_0 V}\left|\frac{\partial U}{\partial x}\right| \tag{43}$$

normal regions:

$$\alpha^2=-\frac{k^2}{\rho_0 V}\frac{\gamma p}{\rho_0}. \tag{44}$$

It is seen that small disturbances are damped out in the shock layers but propagate without either growth or decay in normal regions. This corresponds to physical reality, so our expression (7) is satisfactory also as regards stability of the resulting differential equations.

We can furthermore identify the terms in the equations of variation that lead to the dominant terms in (42). They give:

shock regions:

$$\frac{\partial}{\partial t}\delta U\approx\sigma\frac{\partial^2}{\partial x^2}\delta U, \tag{45}$$

where

$$\sigma=\frac{2(c\Delta x)^2}{V\rho_0}\left|\frac{\partial U}{\partial x}\right|;$$

normal regions:

$$\frac{\partial^2}{\partial t^2}\delta U\approx s_0^2\frac{\partial^2}{\partial x^2}\delta U, \tag{46}$$

where

$$s_0^2=\frac{\gamma p}{\rho_0^2 V}.$$

Thus our system has the character of a diffusion equation in the shock layers and of a sound equation elsewhere.

VI. FINITE DIFFERENCE EQUATIONS

There are many systems of finite difference equations equivalent to the differential equations, but we shall restrict the discussion to one of the simplest. Certain other systems are much superior from the point of view of stability but are more complicated from the point of view of numerical solution. Systems of the latter type will be discussed elsewhere.

Let the points of a rectangular network with spacings Δx and Δt be denoted by x_l, $t^n (l=0, 1, 2, \cdots, L; n=0, 1, 2, \cdots)$. We shall also have occasion to deal with intermediate points, having coordinates $x_{l+\frac{1}{2}} \overset{\text{def.}}{=} \frac{1}{2}(x_{l+1}+x_l)$ etc. To facilitate the writing, we introduce abbreviations such as:

$$V_{l+\frac{1}{2}}{}^n = V(x_{l+\frac{1}{2}}, t^n) \text{ etc.} \tag{47}$$

The difference equations corresponding to (3), (5), (8)' and (36) are:

$$\rho_0 \frac{U_l{}^{n+\frac{1}{2}} - U_l{}^{n-\frac{1}{2}}}{\Delta t} = -\frac{p_{l+\frac{1}{2}}{}^n + q_{l+\frac{1}{2}}{}^{n-\frac{1}{2}} - p_{l-\frac{1}{2}}{}^n - q_{l-\frac{1}{2}}{}^{n-\frac{1}{2}}}{\Delta x}, \tag{48}$$

$$\rho_0 \frac{V_{l+\frac{1}{2}}{}^{n+1} - V_{l+\frac{1}{2}}{}^n}{\Delta t} = \frac{U_{l+1}{}^{n+\frac{1}{2}} - U_l{}^{n+\frac{1}{2}}}{\Delta x}, \tag{49}$$

$$q_{l+\frac{1}{2}}{}^{n+\frac{1}{2}} = -\frac{2(c\Delta x)^2}{V_{l+\frac{1}{2}}{}^n + V_{l+\frac{1}{2}}{}^{n+1}} \frac{(U_{l+1}{}^{n+\frac{1}{2}} - U_l{}^{n+\frac{1}{2}}) \cdot |U_{l+1}{}^{n+\frac{1}{2}} - U_l{}^{n+\frac{1}{2}}|}{(\Delta x)^2}, \tag{50}$$

and

$$\left[\gamma \frac{p_{l+\frac{1}{2}}{}^{n+1} + p_{l+\frac{1}{2}}{}^n}{2} + (\gamma-1) q_{l+\frac{1}{2}}{}^{n+\frac{1}{2}}\right] \frac{V_{l+\frac{1}{2}}{}^{n+1} - V_{l+\frac{1}{2}}{}^n}{\Delta t} + \frac{V_{l+\frac{1}{2}}{}^{n+1} + V_{l+\frac{1}{2}}{}^n}{2} \frac{p_{l+\frac{1}{2}}{}^{n+1} - p_{l+\frac{1}{2}}{}^n}{\Delta t} = 0. \tag{51}$$

These equations are correct to second order of small quantities Δx and Δt, except for the terms containing q in (48) which are only correct to the first order; but these terms are negligible except in the shock layers and are physically artificial in any case.

For numerical solution, suppose that all quantities are known for superscript n or less. Compute $U_l{}^{n+\frac{1}{2}}$ from (48) for each l; compute $V_{l+\frac{1}{2}}{}^{n+1}$ from (49) for each l; compute $q_{l+\frac{1}{2}}{}^{n+\frac{1}{2}}$ from (50) for each l; compute $p_{l+\frac{1}{2}}{}^{n+\frac{1}{2}}$ from (51) for each l; this completes a cycle. Boundary conditions are needed, an example being (rigid walls at ends of a tube): $U_0{}^{n+\frac{1}{2}} \equiv 0$, $U_L{}^{n+\frac{1}{2}} \equiv 0$.

VII. STABILITY OF THE DIFFERENCE EQUATIONS

Equations (48) to (51), being only approximations to the differential equations, cannot be expected to give all features of the solution with precision. If U, V are thought of as being expanded in Fourier series with coefficients depending on t, the long-wave-length components are accurately given by (48) to (51), provided Δx and Δt are sufficiently small, but the components whose wave-lengths are of order Δx are always falsified somewhat. This falsification is harmless, provided that all physically relevant components are treated accurately; this requires not only that Δx and Δt be small but also that the physically insignificant components with wave-lengths of order Δx remain small during the entire calculation. It may happen not only that the short-wave-length components are falsified but also that their amplitudes increase with increasing n, in spite of the stability of the *differential* equations. This increase is generally exponential and, if it occurs, quickly makes gibberish of the entire calculation.

The avoidance of such catastrophes, when partial differential equations are approximated by difference equations, has been the subject of study by various investigators, beginning with the fundamental paper of Courant, Friedrichs, and Lewy referred to in reference 1. We shall give a somewhat heuristic discussion of the stability questions met in the present problem.

We again suppose a small perturbation δU, δV, etc., superposed on a smooth solution, and consider the equations of variation of (48) to (51). According to the analysis of Part V, the dominant terms are expected to be (see Eqs. (45) and (46)):

shock regions:

$$\frac{\delta U_l{}^{n+\frac{1}{2}} - \delta U_l{}^{n-\frac{1}{2}}}{\Delta t} \approx \sigma \frac{\delta U_{l+1}{}^{n-\frac{1}{2}} - 2\delta U_l{}^{n-\frac{1}{2}} + \delta U_{l-1}{}^{n-\frac{1}{2}}}{(\Delta x)^2}; \tag{52}$$

normal regions:

$$\frac{\delta U_l{}^{n+\frac{1}{2}} - 2\delta U_l{}^{n-\frac{1}{2}} + \delta U_l{}^{n-\frac{1}{2}}}{(\Delta t)^2} \approx s_0{}^2 \frac{\delta U_{l+1}{}^{n-\frac{1}{2}} - 2\delta U_l{}^{n-\frac{1}{2}} + \delta U_{l-1}{}^{n-\frac{1}{2}}}{(\Delta x)^2}; \tag{53}$$

and this can be verified by writing out the difference equations of variation in detail.

As before, we consider perturbations of the form:

$$\delta U = \delta U_0 e^{ikx+\alpha t}, \text{ etc.}, \tag{54}$$

so that:

$$\delta U_l{}^{n+\frac{1}{2}} = \delta U_0 \zeta^l \xi^{n+\frac{1}{2}}, \tag{55}$$

where:

$$\zeta = e^{ik\Delta x}, \quad \xi = e^{\alpha \Delta t}. \tag{56}$$

For stability we require that $|\xi| \leq 1$ for all real k.

We consider normal regions first. Substitution of (55) into (53) and cancellation of common factors gives:

$$\xi - 2 + \frac{1}{\xi} = 2\left(\frac{s_0 \Delta t}{\Delta x}\right)^2 (\cos k\Delta x - 1). \tag{57}$$

If we call

$$L=\frac{s_0\Delta t}{\Delta x}, \tag{58}$$

the solution of (57) is

$$\xi_{1,2}=b\pm(b^2-1)^{\frac{1}{2}} \tag{59}$$

where

$$b=1-L^2(1-\cos k\Delta x). \tag{60}$$

According to (60), b is always <1, so there are two cases:

$$-1<b<1, \quad |\xi_1|=|\xi_2|=1; \quad \text{stability;} \tag{61}$$

$$b<-1, \quad |\xi_1|<1<|\xi_2|; \quad \text{instability.} \tag{62}$$

From Eq. (60) it is seen that (61) will hold for all k if and only if

$$L\leq 1. \tag{63}$$

This is the familiar condition for stability of hydrodynamical equations of the form (53).

A similar treatment of Eq. (52) for the shock region yields directly:

$$\xi-1=2\frac{\sigma\Delta t}{(\Delta x)^2}(\cos k\Delta x-1), \tag{64}$$

and the stability condition is clearly that

$$\frac{2\sigma\Delta t}{(\Delta x)^2}\leq 1 \quad \text{or} \quad \Delta t\leq\frac{(\Delta x)^2}{2\sigma}. \tag{65}$$

To interpret this result, we calculate σ according to (45) for the steady shock discussed in Part IV. From (33), (34);

$$\left|\frac{\partial U}{\partial x}\right|=\rho_0\left|\frac{\partial V}{\partial t}\right|=\rho_0 s\left|\frac{\partial V}{\partial w}\right|$$

$$=\rho_0 s\frac{V_i-V_f}{2c\Delta x}((\gamma+1)/2)^{\frac{1}{2}}\cos\frac{w}{w_0}. \tag{66}$$

Therefore

$$\sigma=sc\Delta x\frac{V_i-V_f}{V}((\gamma+1)/2)^{\frac{1}{2}}\cos\frac{w}{w_0}, \tag{67}$$

and the stability condition (65) takes the form:

$$\Delta t\leq\frac{V\Delta x}{2sc(V_i-V_f)}\left(\frac{2}{\gamma+1}\right)^{\frac{1}{2}}\frac{1}{\cos(w/w_0)}, \tag{68}$$

or,

$$\Delta t\leq\frac{\Delta x}{4sc}\left(\frac{2}{\gamma+1}\right)^{\frac{1}{2}}\frac{(\eta+1)/(\eta-1)+\sin(w/w_0)}{\cos(w/w_0)}, \tag{69}$$

by further use of (34), where

$$\eta=V_i/V_f. \tag{70}$$

The quantity η is a measure of the shock strength. Equation (69) shows that different parts of the shock layer (i.e., different values of w) impose different restrictions on Δt. The effective restriction is obtained by replacing the right member of (69) by its minimum value for $-(\pi/2)w_0\leq w\leq(\pi/2)w_0$. The minimum of the last factor in (69) is found to be $2(\eta)^{\frac{1}{2}}/(\eta-1)$, and the stability condition is

$$\Delta t\leq\frac{\Delta x}{2sc}\left(\frac{2}{\gamma+1}\right)^{\frac{1}{2}}\frac{\eta^{\frac{1}{2}}}{\eta-1}. \tag{71}$$

For practical application, it is convenient to express (71) in terms of the quantity L appearing in the normal hydrodynamic stability condition (63). By elimination of p_i from the Hugoniot Eqs. (21), (22), and (8), the speed of the shock is found to be

$$\left.\begin{aligned} s=\frac{M}{\rho_0}&=\frac{1}{\rho_0}\left(\frac{2}{(\gamma+1)\eta-(\gamma-1)}\right)^{\frac{1}{2}}\left(\gamma\frac{p_f}{V_f}\right)^{\frac{1}{2}}\\ &=\left(\frac{2}{(\gamma+1)\eta-(\gamma-1)}\right)^{\frac{1}{2}}s_{0f},\end{aligned}\right\} \tag{72}$$

where s_{0f} is the speed of sound, relative to Lagrangean coordinates, in the material behind the shock. By combining (71) and (72) and use of (58), the stability condition becomes:

$$L_f=\frac{s_{0f}\Delta t}{\Delta x}\leq\frac{1}{2c}\frac{\{[\eta-(\gamma-1)/(\gamma+1)]\eta\}^{\frac{1}{2}}}{\eta-1}. \tag{73}$$

Lastly, the shock strength, and hence η, is generally unknown until the calculation has been performed; it is therefore advisable to replace the right member of (73) by its minimum with respect to η. η can vary in the range $1\leq\eta\leq(\gamma+1)/(\gamma-1)$ (the latter value corresponding to an infinitely strong shock) and the minimum of (73) is attained at the upper end of this range, where the last factor in (73) has the value $\gamma^{\frac{1}{2}}$. Our final, sufficient condition for stability reads:

$$L_f\leq\gamma^{\frac{1}{2}}/2c. \tag{74}$$

This condition has been found to insure stability in test calculations, whereas a serious violation of it leads to trouble. The choice $c=1$ has been found to yield good results in practice for the representation of shocks, in which case the stability condition is at worst slightly more severe than the one that must be observed, anyway (compare Eq. (63)), to insure stability of the motion behind the shock.

Some Non-Linear Systems Permitting Simple Harmonic Motion

JOHN E. BROCK
Department of Mechanics, Washington University, St. Louis, Missouri
(Received October 24, 1949)

By elementary and obvious methods a differential equation is obtained, which in general is non-linear and usually affords a solution which is a simple sinusoidal function of the independent variable (i.e., simple harmonic motion). Two particular examples are studied briefly although no physical realizations are presented. One is led to speculate that new types of oscillators having a sinusoidal output are described by these equations, and that practical applications, such as to frequency metering, may eventually be evolved. However, the writer's speculations along these lines have as yet borne no fruit and his primary object in publishing this material is to bring to the attention of those interested in non-linear phenomena, about which our understanding is so essentially imperfect, an analysis which, though itself possibly trivial, conceivably may prove of service in setting someone else off on a productive investigation.

THE ordinary linear oscillator described by the equation

$$\ddot{x}+\omega^2 x=0, \tag{1}$$

in which $\omega^2>0$, permits simple harmonic motion given by

$$x=A\sin(\omega t+\phi), \tag{2}$$

in which A and ϕ are arbitrary constants, determined by the initial conditions of the motion and independent of ω^2.

This is a special case of the more general oscillator described by

$$F[-(\ddot{x}/x),\quad x^2-(x\dot{x}^2/\ddot{x})]=0, \tag{3}$$

in which F represents an arbitrary function of two variables. Equation (3) possesses the solution given by Eq. (2) subject to the condition

$$F(\omega^2,A^2)=0. \tag{4}$$

That is, frequency ω and amplitude A are related as indicated by Eq. (4). Equation (3) may also possess solutions other than that given by Eq. (2).

Equation (3) is derived as follows. From Eq. (2) we also have

$$\dot{x}=A\omega\cos(\omega t+\phi), \tag{5}$$

$$\ddot{x}=-A\omega^2\sin(\omega t+\phi). \tag{6}$$

Eliminating A between Eqs. (2) and (6), we get

$$\omega^2=-\ddot{x}/x, \tag{7}$$

and eliminating ω between Eqs. (2), (5), and (6), we get

$$A^2=x^2-x\dot{x}^2/\ddot{x}. \tag{8}$$

If the solution, Eq. (2), is to be subject to the condition, Eq. (4), the desired differential equation clearly is given by Eq. (3).

The solutions to Eq. (3) given by Eq. (2) and subject to Eq. (4) represent conservative motions, for writing

$$T=\tfrac{1}{2}\dot{x}^2, \tag{9}$$

$$V=-\tfrac{1}{2}x\ddot{x}=\tfrac{1}{2}x^2\left(-\frac{\ddot{x}}{x}\right)=\tfrac{1}{2}\omega^2x^2, \tag{10}$$

we have, from Eq. (3),

$$F\left(\omega^2,\frac{T+V}{\tfrac{1}{2}\omega^2}\right)=0, \tag{11}$$

and, from Eq. (4), this yields

$$T+V=\tfrac{1}{2}\omega^2A^2=\text{const.} \tag{12}$$

From Eq. (10), it is seen that the systems under consideration involve what is essentially a non-linear spring in which the effective spring constant is ω^2, and thus generally a function of amplitude.

If the implicit function $F(\omega^2,A^2)=0$ may be written explicitly as

$$\omega^2=f(A^2) \tag{13}$$

or

$$A^2=g(\omega^2), \tag{14}$$

Eq. (3) takes the form

$$\ddot{x}+xf[x^2-(x\dot{x}^2/\ddot{x})]=0 \tag{15}$$

or

$$(x\dot{x}^2/x)-x^2+g(-\ddot{x}/x)=0, \tag{16}$$

as the case may be.

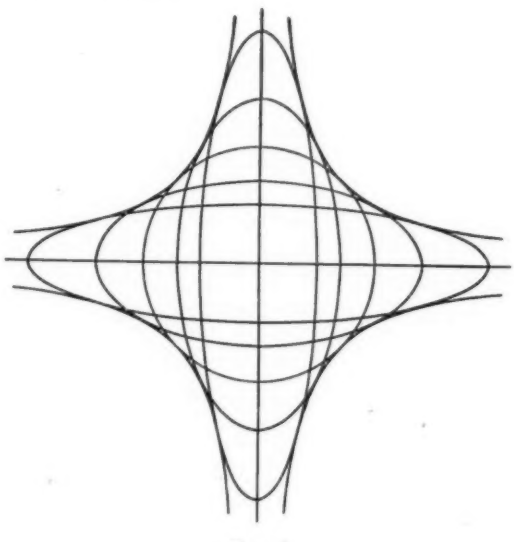

FIG. 1.

Defining the new function

$$h(z)=zg(z) \tag{17}$$

permits writing Eq. (16) in the more convenient form

$$x\ddot{x}-\dot{x}^2-h(\ddot{x}/x)=0. \tag{18}$$

From Eqs. (13) and (14), it is observed that f and g are non-negative functions which need be defined only for positive values of the argument. Thus, these remarks are also true of the function h.

Equation (3) as written may be of degree greater than one, and this leads to ambiguities in interpreting the physical motion. It is, in effect, equivalent to the product of one or more equations of the type

$$\ddot{x}+x\psi_i(x,\dot{x})=0 \tag{19}$$

each of first degree and capable of physical interpretation.

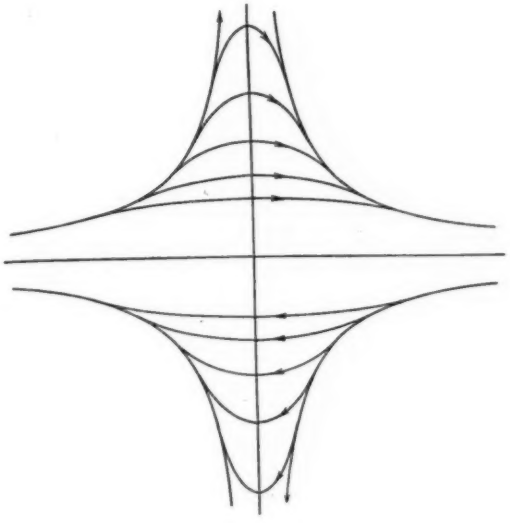

FIG. 2.

From Eq. (19), it is seen that

$$\psi=-\ddot{x}/x \tag{20}$$

and

$$x^2+(\dot{x}^2/\psi)=x^2-(x\dot{x}^2/x). \tag{21}$$

Thus,

$$F[\psi,x^2+(\dot{x}^2/\psi)]=0 \tag{22}$$

can be solved for

$$\psi=\psi_i(x,\dot{x}), \tag{23}$$

where the subscript $_i$ indicates that there may be several solutions.

Further points of interest will be forthcoming from examination of two simple examples. For convenience, the possible physical motions are discussed by reference to their trajectories in the displacement-velocity

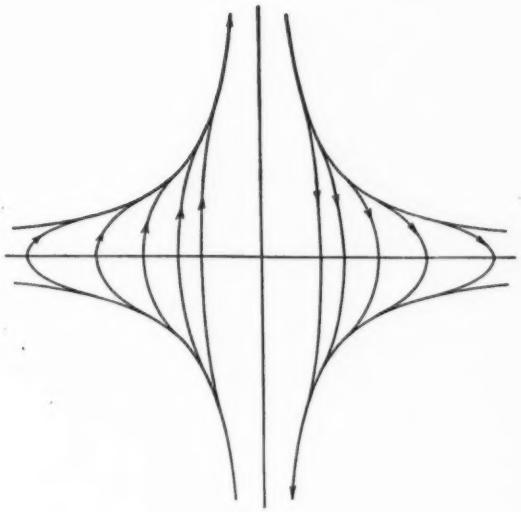

FIG. 3.

"phase" plane.[1] In these discussions, the notation

$$y=\dot{x} \tag{24}$$

will be used.

EXAMPLE 1

Let the relation between frequency and amplitude be given by

$$\omega^2=M^2/A^4, \quad (M>0). \tag{25}$$

Equation (3) becomes

$$\frac{\ddot{x}}{x}+\frac{M^2}{(x^2-x\dot{x}^2/\ddot{x})^2}=0, \tag{26}$$

with a solution

$$x=A\sin[(Mt/A^2)+\phi] \tag{27}$$

for which the phase trajectories are given by

$$(x^2/A^2)+(A^2y^2/M^2)=1. \tag{28}$$

The solution represents simple harmonic motion and the trajectories are concentric ellipses each of area πM. The trajectories are shown in Fig. 1. These ellipses are tangent to the equilateral hyperbolas

$$4x^2y^2=M^2 \tag{29}$$

at the points

$$\left.\begin{matrix}x=\pm(A\sqrt{2})/2\\ y=\pm(M\sqrt{2})/2A\end{matrix}\right\}. \tag{30}$$

Through every point for which $4x^2y^2<M^2$ there pass two ellipses and there are obvious difficulties in attempting to interpret the physical motion.

---

[1] For a discussion of and examples illustrating this type of representation, cf., N. Minorski, *Introduction to Non-Linear Mechanics* (J. W. Edwards, Ann Arbor, Michigan, 1947), especially Chapters 1–3.

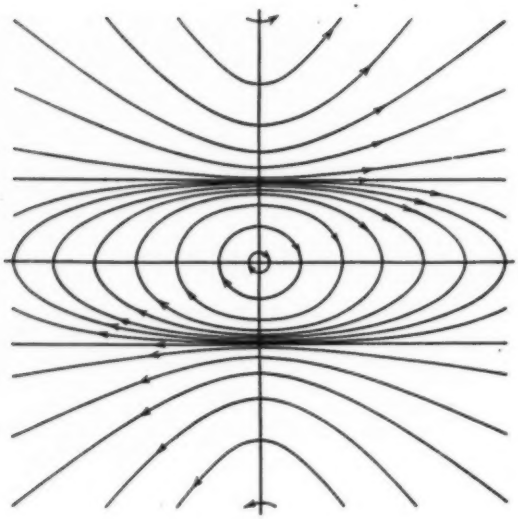

FIG. 4.

In this case, Eq. (22) becomes

$$\psi=\frac{M^2}{(x^2+y^2/\psi)^2} \tag{31}$$

from which we obtain the following two equations in the form of Eq. (19):

$$\ddot{x}+\frac{1}{x^3}\left\{\frac{M^2}{2}-x^2\dot{x}^2-M\left(\frac{M^2}{4}-x^2\dot{x}^2\right)^{\frac{1}{2}}\right\}=0 \tag{32}$$

$$\ddot{x}+\frac{1}{x^3}\left\{\frac{M^2}{2}-x^2\dot{x}^2+M\left(\frac{M^2}{4}-x^2\dot{x}^2\right)^{\frac{1}{2}}\right\}=0. \tag{33}$$

These two equations are, of course, also obtained by solving Eq. (26) for $\ddot{x}$. In each case the accelerations are real only if

$$4x^2y^2\leqq M^2, \tag{34}$$

which explains why no trajectories are found outside the region bounded by the hyperbolas of Eq. (29).

In the case of Eq. (32), a solution is given by Eq. (27), and this solution is valid only for

$$x^2\leqq A^2/2, \tag{35}$$

in other words, only for those values of x between the points where the elliptical trajectory (given by Eq. (28)) becomes tangent to the hyperbolas of Eq. (29), (cf. Eq. (30)).

In the case of Eq. (33), the solution afforded by Eq. (27) is valid only for

$$x^2\geqq A^2/2. \tag{36}$$

In both cases, the hyperbolic envelope also represents a solution valid for all x, the solution being

$$x^2=C\pm Mt, \tag{37}$$

the corresponding trajectories being the hyperbolas themselves (Eq. (29)).

Figures 2 and 3 show the trajectories corresponding to Eq. (32) and Eq. (33), respectively. It is seen that Fig. 1 represents a superposition of these.

In the cases of Eq. (32), by writing

$$\begin{cases} y=\dot{x}=F_1(t;x,y) \\ \dot{y}=-(1/x^3)\{(M^2/2)-x^2y^2 \\ \qquad -M[(M^2/4)-x^2y^2]^{\frac{1}{2}}\}=F_2(t;x,y) \end{cases} \tag{38}$$

we have a system of two first-order equations to which the ordinary test for uniqueness applies.[2] It is clear that both functions F_1 and F_2 satisfy Lipschitz conditions

$$|F_i(t;x_1,y_1)-F_i(t;x_2,y_2)|<k(|x_1-x_2|+|y_1-y_2|) \tag{39}$$

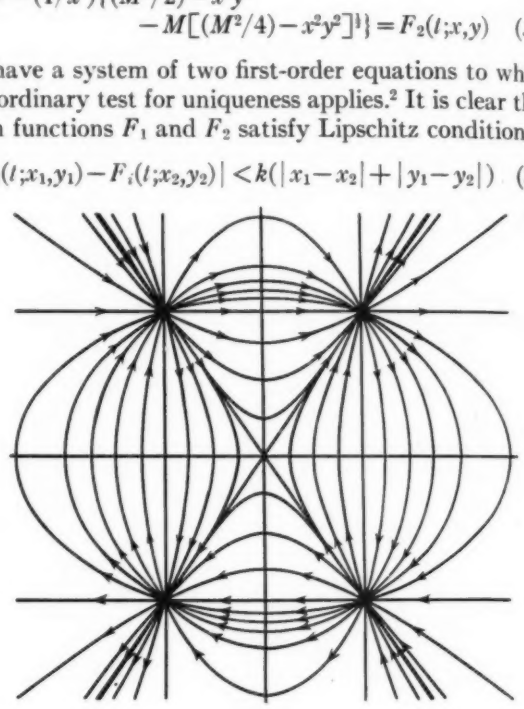

FIG. 5.

in any domain D which does not include any part of the y axis or any part of the hyperbolas of Eq. (29). Since (real) solutions are defined only for $4x^2y^2\leqq M^2$, we can assert their uniqueness except possibly for $x=0$, and upon the hyperbolas.

The trajectories of Fig. 2 show that indeed the solution is not unique on the hyperbolas. However, a unique solution does exist for any set of initial conditions corresponding to a point on the interior of the region bounded by the hyperbolas. For representative points on the y-axis, the situation must be investigated by the following limiting process:

$$\begin{aligned}\operatorname*{Lim}_{x\to 0}\ddot{x}&=\operatorname*{Lim}_{x\to 0}\left[-\frac{1}{x^3}\left\{(M^2/2)-x^2y^2 \right.\right. \\ &\qquad \left.\left. -M[(M^2/4)-x^2y^2]^{\frac{1}{2}}\right\}\right] \\ &=\operatorname*{Lim}_{x\to 0}\left[-(xy^4/M^2)+0(x^3y^6/M^4)\right]=0.\end{aligned} \tag{40}$$

(This result holds regardless of the value of y.) Thus, a particle starting with any velocity and zero displace-

[2] Cf. E. L. Ince, *Ordinary Differential Equations* (Dover Publications, New York, 1944), p. 71.

ment continues with the same velocity until the representative point moves off the y-axis into the region to which the preceding uniqueness test applies. The origin provides an exception to this statement, but it is clear that in this case the particle will remain with zero velocity and zero displacement.

The trajectories of Fig. 2 do not clearly show the behavior if the initial conditions involve $y_0=0$. In such a case, the acceleration vanishes identically and the solutions is $x\equiv x_0$; such a state of equilibrium is evidently unstable since if any positive (negative) velocity is given to the particle, its displacement will tend to positive (negative) infinity.

A somewhat similar analysis may be applied to Eq. (33). This will not be given here in full detail, but it may be remarked that the motion is not defined if the initial conditions involve zero displacement, i.e., the representative point on the y-axis.

Subject to the exceptions noted in the two preceding paragraphs, it is seen that, if the initial conditions are such that

$$4x_0^2\dot{x}_0^2<M^2, \tag{41}$$

the motion will be simple harmonic until the representative point in the phase plane, moving along an elliptical path, touches one of the bounding hyperbolas. From this time on, the point follows the hyperbola and the motion is described by Eq. (37). There is no ambiguity unless the representative point for the initial conditions lies upon the bounding hyperbola in either the second or fourth quadrants. In such a case there is an instability and the representative point may follow a hyperbola or break away and follow an ellipse at any time. It may be remarked that hyperbolic trajectories have some properties similar to those of "limit cycles."

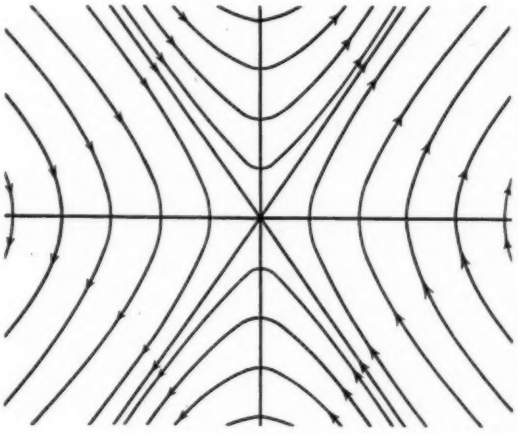

FIG. 6.

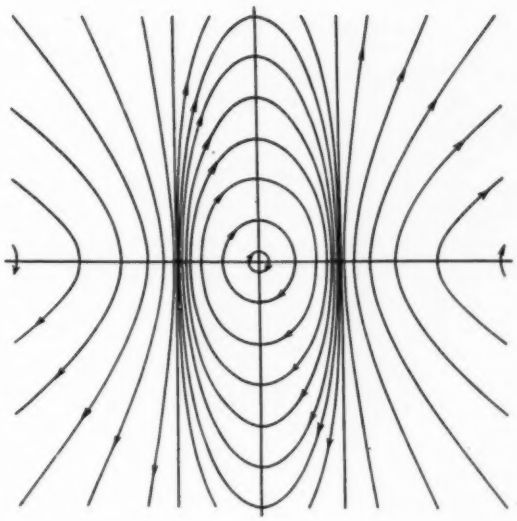

FIG. 7.

EXAMPLE 2

A fairly general case is represented by the condition

$$\omega^2=\frac{aA^2+b}{cA^2+d}, \tag{42}$$

where a, b, c, and d are constants such that $ad\neq bc$. We are led to the equations

$$\ddot{x}+x\left\{\frac{(ax^2+b-c\dot{x}^2)\pm[(ax^2+b-c\dot{x}^2)^2+4a\dot{x}^2(cx^2+d)]^{\frac{1}{2}}}{2(cx^2+d)}\right\}. \tag{43}$$

The function

$$x=A\sin\left[\left(\frac{aA^2+b}{cA^2+d}\right)^{\frac{1}{2}}t+\omega\right] \tag{44}$$

is a solution to Eq. (43) with the plus sign if and only if

$$\frac{acA^4+[bc\sin^2\theta+ad(1+\cos^2\theta)]A^2+bd}{(cA^2+d)}<0, \tag{45}$$

where

$$\theta=\left(\frac{aA^2+b}{cA^2+d}\right)^{\frac{1}{2}}t+\phi. \tag{46}$$

Similarly, Eq. (44) is a solution to Eq. (43) with the minus sign if and only if the inequality opposite to that of condition (45) is satisfied.

This example is too involved to work out in full detail here, so we consider the special case where $a=0$. With this simplification we are led to:

$$[x^2+(d/c)]\ddot{x}/x=\dot{x}^2-(b/c). \tag{47}$$

In this case the investigation naturally divides itself into four cases.

Case (1). $(d/c)>0$; $(b/c)>0$

If $\dot{x}^2<(b/c)$, a solution is

$$x=A\sin\left[\left(\frac{b}{cA^2+d}\right)^{\frac{1}{2}}t+\phi\right] \tag{48}$$

with trajectories

$$x^2/A^2+[(cA^2+d)/(bA^2)]y^2=1. \tag{49}$$

If $\dot{x}^2>(b/c)$, a solution is

$$x=A\sinh\left[\left(\frac{b}{cA^2-d}\right)^{\frac{1}{2}}t+\phi\right] \tag{50}$$

with trajectories

$$[y^2(cA^2-d)/bA^2]-(x^2/A^2)=1. \tag{51}$$

For the whole phase plane the trajectories are shown in Fig. 4.

Case (2). $(d/c)<0$; $(b/c)>0$

In this case if $A^2<-(d/c)$, the frequency becomes imaginary and we cannot expect simple harmonic motion. The solution and trajectories given by Eqs. (48) and (49) hold for $-(d/c)<A^2<\infty$. For $A^2<-(d/c)$, the solutions and trajectories of Eqs. (50) and (51) hold as well as those given by

$$x=A\cosh\left[\left(\frac{-b}{cA^2+d}\right)^{\frac{1}{2}}t+\phi\right] \tag{52}$$

and

$$(x^2/A^2)+(y^2(cA^2+d)/A^2b)=1. \tag{53}$$

All the trajectories pass through the points $x=\pm(-d/c)^{\frac{1}{2}}$, $y=\pm(b/c)^{\frac{1}{2}}$, which represent instability in the sense that many motions of the representative point are possible at these points. The trajectories are shown in Fig. 5.

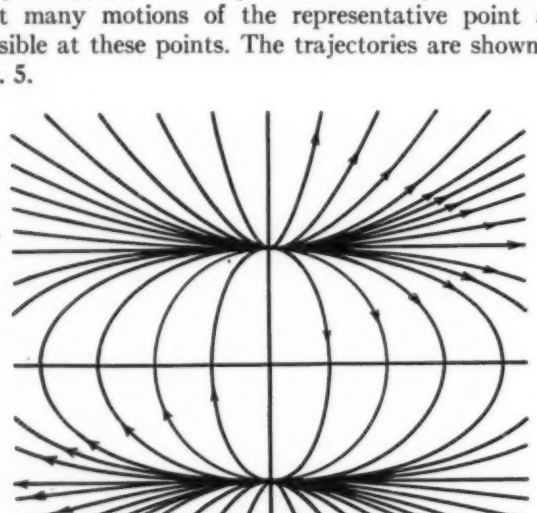

FIG. 8.

Case (3). $(d/c)>0$; $(b/c)<0$

In this case the frequency is imaginary and there can be no periodic motions as may be seen directly from Eq. (47). The singularity at $x=\dot{x}=0$ is an ordinary saddle point. The solutions and corresponding trajectories are given by Eqs. (50), (51), (52), and (53). The trajectories are shown in Fig. 6.

Case (4). $(d/c)<0$; $(b/c)<0$

If $A^2<-(d/c)$, Eqs. (48) and (49) give the solution and trajectories respectively. If $A^2>-(d/c)$, Eqs. (52) and (53) give these results. The trajectories are shown in Fig. 7.

Two interesting special cases are obtainable from this analysis. If $b=c$, and $d=0$, we have the equation

$$\dot{x}^2-x\ddot{x}=1 \tag{54}$$

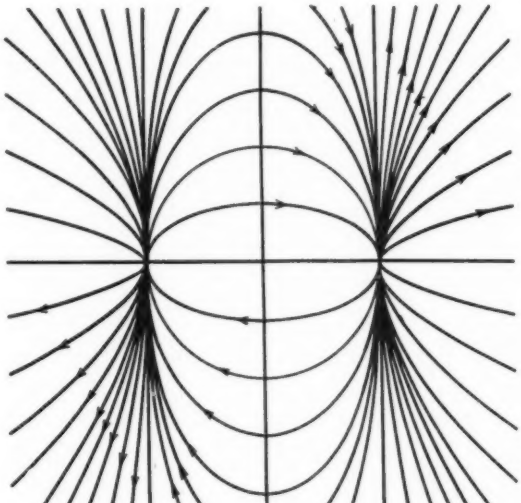

FIG. 9.

for which, if $\dot{x}^2\leqq 1$, the solution is

$$x=A\sin[(t/A)+\phi] \tag{55}$$

and the trajectories are

$$(x^2/A^2)+y^2=1. \tag{56}$$

If $\dot{x}^2\geqq 1$, the solution is

$$x=A\sinh[(t/A)+\phi] \tag{57}$$

and the trajectories are

$$y^2-(x^2/A^2)=1. \tag{58}$$

All trajectories pass through the points $x=0$; $y=\pm 1$. (Fig. 8).

If $b=0$, and $d=-c$, we have the equation

$$\ddot{x}(1-x^2)+x\dot{x}^2=0. \tag{59}$$

If $x^2 \leqq 1$, the solution is given by

$$x = \sin[(t/A)+\phi] \tag{60}$$

and the trajectories by

$$x^2 + A^2y^2 = 1. \tag{61}$$

If $x^2 \geqq 1$, the solution is given by

$$x = \cosh[(t/A)+\phi] \tag{62}$$

and the trajectories by

$$x^2 - A^2y^2 = 1. \tag{63}$$

The trajectories are shown in Fig. 9.

The ordinary Lipschitz condition[2] for (existence and) uniqueness of the solutions suffices to demonstrate uniqueness for solutions to Eq. (47) except for $x = \pm(-d/c)^{\frac{1}{2}}$. The implications of this are apparent in Cases (2) and (4). In Cases (1) and (3) there is no exception to uniqueness.

The Analogy between Hydraulic Jumps in Liquids and Shock Waves in Gases*

F. R. Gilmore, M. S. Plesset, and H. E. Crossley, Jr.
Hydrodynamics Laboratory, California Institute of Technology, Pasadena, California
(Received September 30, 1949)

The theory of the hydraulic jump is presented briefly, and the analogy between this phenomenon and the compression shock wave in gases is pointed out. The results of experimental measurements of hydraulic-jump intersections on a water table are reported. Considerable disagreement between theory and experiment is found. Other investigators have noted a disagreement between theory and experiment for compression-shock intersections in gases. The discrepancy in the aerodynamic case appears unlike that found in the hydraulic case. Possible reasons for the discrepancy in the hydraulic case are discussed; some sources of error are peculiar to hydraulic jumps and do not apply to compression shocks. Such factors limit the utility of the water table as an analog device.

INTRODUCTION

It has been known for a long time that liquid flow in open channels is of two different types: relatively smooth "streaming" flow at low speeds, and high speed "shooting" flow characterized by standing waves and frequently by sudden changes in depth known as hydraulic jumps. More recently, the increased study of compressible-flow phenomena has led investigators to note the resemblances between streaming channel flow and subsonic compressible flow, and between shooting channel flow and supersonic compressible flow. This similarity was put on a mathematical basis by Jouguet[1] and Riabouchinsky.[2] Further theoretical investigations have been carried out by von Kármán[3] and by Preiswerk[4]; experimental studies have been made by Ippen[5] and by Binnie and Hooker.[6] Later work[7] has been concerned with the application of the theory to model testing, except for a recent theoretical paper by Stoker.[8]

The analogy between liquid flow with a free surface and two-dimensional gas flow has been found useful for qualitative investigations of high speed gas flow by means of relatively simple water-table installations. However, recent advances in aerodynamics have brought many problems to the point where more accurate information is required than can be obtained from the hydraulic analogy. The present utility of the water table as an analog device is limited to special problems in aerodynamics, such as transient phenomena in high speed flow, and shock-wave intersection phenomena, particularly of the Mach type.

The discussion here is concerned with the analogy between hydraulic-jump intersections on a water table and shock-wave intersections in gases.

THEORY

Shock Waves in Compressible Gas Flow

As is well known, a compression wave of finite amplitude in a gas grows continually steeper as the wave progresses until a finite limiting steepness determined by viscous and heat-conduction effects is attained. The limiting thickness of the shock in the usual situation is negligible compared with the other dimensions in the flow field. The flow relations across a straight compression shock are determined by the conservation laws for mass, momentum, and energy without consideration of the conditions inside the shock. The results are the

* This study was supported by the ONR and the Navy Bureau of Ordnance.

[1] E. Jouguet, J. de math. pures et appliq. (Series 8) **3**, 1 (1920).
[2] D. Riabouchinsky, Comptes Rendus **195**, 998 (1932); **199**, 632 (1934); **202**, 1725 (1936).
[3] Th. von Kármán, Zeits. f. angew. Math. Mech. **18**, 49 (1938).
[4] E. Preiswerk, N.A.C.A. Tech. Memo. 934 and 935 (1940); translated from Mitteilungen Inst. f. Aerodyn., No. 7, Eidg. Tech. Hochsch., Zurich (1938).
[5] A. Ippen, Ph.D. thesis, California Institute of Technology (1936).
[6] A. M. Binnie and S. G. Hooker, Proc. Roy. Soc. **159**, 592 (1937).
[7] See, for example, Orlin, Lindner, and Bitterly, N.A.C.A. Report 875 or Tech. Note 1185 (1947).
[8] J. J. Stoker, Commun. App. Math. **1**, 1 (1948).

familiar Rankine-Hugoniot equations.[9] Figure 1 shows curves of the density ratio, temperature ratio, and the square root of the pressure ratio across a normal shock as functions of the initial Mach number for perfect gases having specific-heat ratios of $\gamma=1.4$ and $\gamma=2$.

The flow relations across a straight shock which is oblique to the flow may be obtained by superimposing on the normal velocity an additional velocity tangential to the shock which remains unchanged across the shock. For a curved shock, if the radius of curvature is much larger than the shock thickness, the flow relations may be obtained by approximating the curved shock with a number of straight segments.

Hydraulic Jumps in Liquid Flow with a Free Surface

If an elevation wave of finite amplitude is produced on the surface of a liquid, Stoker[8] has shown from the equations of motion that the wave front will grow continually steeper as it progresses, until finally the wave leans forward. At this point experiments show that the wave form deviates from the theoretical form, with the formation of a breaker or "roller" and the establishment of a turbulent wave of constant shape (except for minor fluctuations of short period). Such a steady finite wave is called a hydraulic jump. Bakhmeteff and Matzke[10] have found experimentally that

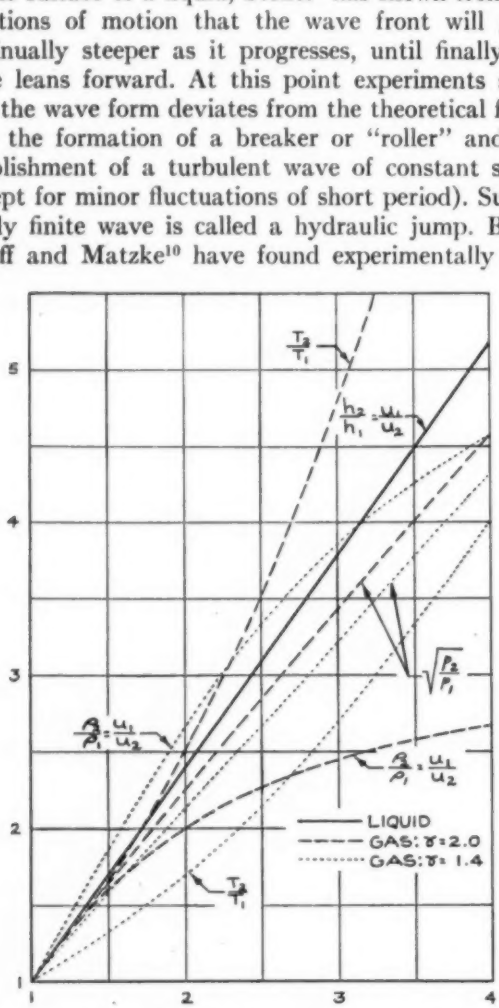

FIG. 1. Flow relations across a normal shock, or hydraulic jump. M_1 = Mach number ahead of shock.

[9] See R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves* (Interscience Publishers, Inc., New York, 1948), pp. 116–125, or similar texts on compressible flow.

[10] B. Bakhmeteff and A. Matzke, Trans. A.S.C.E. **101**, 630 (1936).

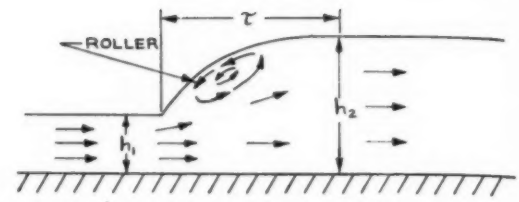

FIG. 2. Typical profile of a strong hydraulic jump ($h_2/h_1>2$).

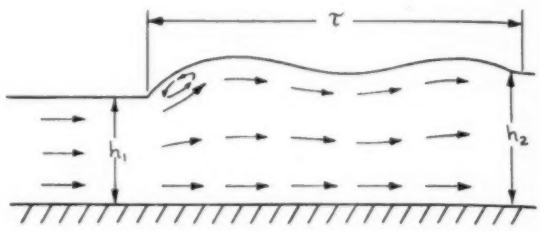

FIG. 3. Typical profile of a weak hydraulic jump ($h_2/h_1<2$).

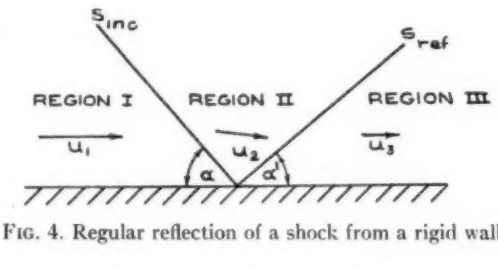

FIG. 4. Regular reflection of a shock from a rigid wall.

strong hydraulic jumps with a height ratio, h_2/h_1, greater than 2 have the simple profile shown in Fig. 2; in these strong hydraulic jumps, the thickness, τ, of the jump is four to six times the height of the jump, h_2-h_1. Weaker jumps generally have the undulatory form shown in Fig. 3. The surface first rises above the final level and then oscillates with diminishing amplitude about the final level. The thickness of the undulatory jump is not well defined, but the undulations generally die out at distances of three to ten times the initial water height, h_1, the distance being greater for the smaller jumps. A theoretical explanation of this general behavior has not as yet been given.

As in the case of compression shocks in gases, flow relations across straight hydraulic jumps are determined by the conservation laws for mass, momentum, and energy without consideration of the conditions inside the jump. The result for the height and velocity ratio of a jump normal to the flow is

$$h_2/h_1=u_1/u_2=(2u_1^2/gh_1+\tfrac{1}{4})^{\frac{1}{2}}-\tfrac{1}{2}. \qquad (1)$$

The height ratio given by Eq. (1) is plotted against initial Mach number as the solid curve in Fig. 1. "Mach number" in this case is defined as $u/(gh)^{\frac{1}{2}}$, since $(gh)^{\frac{1}{2}}$ is the velocity of infinitesimal surface waves.

Flow relations across oblique hydraulic jumps may be calculated in the same way as for oblique compression shocks. Similarly, flow relations across curved jumps can be found, provided the radius of curvature is much larger than the jump thickness. In many experi-

mental situations, however, the thickness of the jump may be of the same order of magnitude as the radius of curvature. Existing theory does not cover this case.

The Analogy between Shocks and Jumps

It may be shown[1–4] that the smooth free-surface flow of a liquid over a horizontal bed obeys equations of motion similar to those for the isentropic two-dimensional flow of a hypothetical gas with specific heat ratio $\gamma=2$. The mathematical correspondence is between the following quantities: the liquid velocity ratio and the gas velocity ratio; and the liquid depth ratio on the one hand and the gas density ratio, the gas temperature ratio, and the square root of the gas pressure ratio, on the other hand. It is to be noted that the isentropic condition makes the gas density ratio, the gas temperature ratio and the square root of the gas pressure ratio all equal when $\gamma=2$.

Flows with shocks or jumps are not isentropic and it might be expected that the quantitative correspondence breaks down. The entropy change, however, is of the third order in the shock strength so that weak shocks are very nearly isentropic; and the quantitative similarity should then hold to a good approximation. This situation is demonstrated in Fig. 1, which compares values of analogous quantities for the compression shock and the hydraulic jump. The curves for a gas with $\gamma=2$ approach the liquid curve asymptotically as the shock becomes weaker.

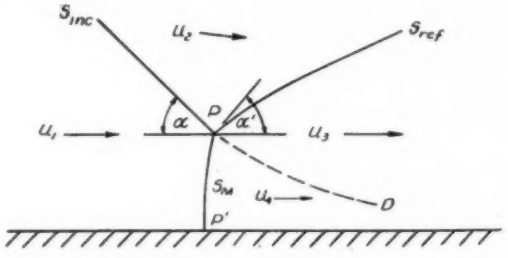

FIG. 5. Mach reflection of a shock from a rigid wall.

An analogy based on a hypothetical gas with $\gamma=2$ is somewhat unrealistic since kinetic theory shows that gases must have specific-heat ratios in the range $1<\gamma\leqslant 5/3$. Figure 1 includes curves for a gas with $\gamma=1.4$, which is the value for air. Although these curves do not have an asymptotic approach to the liquid curve as the shock strength is decreased, some of the flow parameters, e.g., u_1/u_2, have a smaller deviation over a range of strong shocks than the corresponding parameters for $\gamma=2$.

Intersections and Reflections of Hydraulic Jumps and Shock Waves

The following discussion will apply both to hydraulic jumps and compression shock waves, and the term "shock" will be used to signify either phenomenon.

Figure 4 shows the reflection of a straight shock from a plane rigid wall; it could as well refer to the intersection of two equal straight shocks where the wall becomes the plane of symmetry. The fluid in Region I, ahead of the shocks, is assumed to be either at rest or in a state of uniform motion. It is convenient to introduce a coordinate system relative to which the shocks and their intersection are stationary; in this coordinate system let u_1 be the velocity with which the fluid in Region I approaches the incident shock, S_{inc}. As the flow passes through this shock, it is slowed down and deflected to a velocity u_2 in Region II. The reflected shock, S_{ref}, must deflect the flow from the velocity u_2 to a velocity u_3 in Region III which, like u_1, must be parallel to the boundary wall or to the plane of symmetry. This condition is sufficient to determine the angle and strength of the reflected shock.[11]

When the incident shock is too strong, and when it makes too large an angle α with the wall, there is no real solution for the angle and strength of the reflected

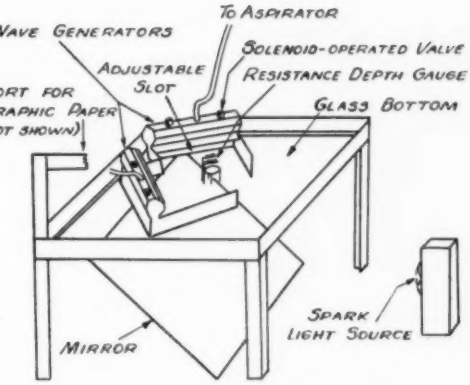

FIG. 6. A schematic representation of the experimental apparatus.

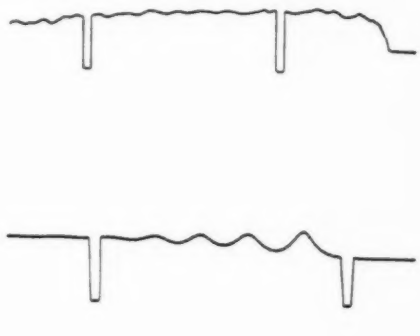

FIG. 7. Typical hydraulic-jump forms. Top curve: Strong jump with irregular roughness. Center curve: Weak jump with undulations. Bottom curve: Smooth jump of medium strength. The breaks in these records, which give the reference zero for depth, are spaced 0.6 sec. apart.

[11] See reference 9, pp. 318–350.

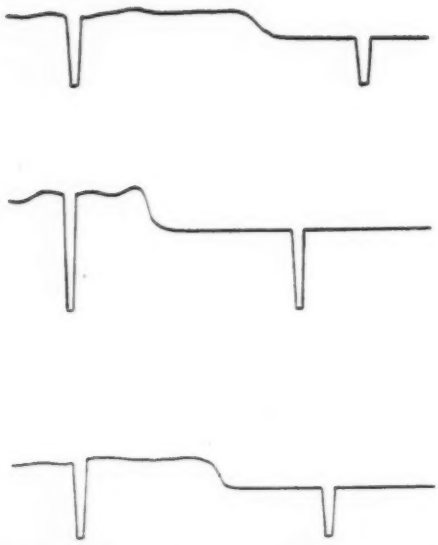

FIG. 8. Effects of surface-tension reducing agents on the jump form. Top curve: Without detergent. Center curve: Strong detergent (lauryl isoquinolinium bromide 0.1 percent). Bottom curve: Weak detergent (Kodak Photo-Flow, 0.5 percent).

shock so that the "regular" reflection scheme shown in Fig. 4 becomes impossible. In this situation the actual flow is found experimentally to follow the scheme shown in Fig. 5. A third shock, the "Mach," S_M, appears, and its length, i.e., the distance PP', increases with time at a fairly constant rate. The Mach, S_M, is frequently curved, and also the reflected shock, S_{ref}, is often curved near the triple-shock intersection.

The simplest theory of the Mach reflection assumes that all three shocks are straight.[11] Coordinates are chosen fixed with respect to the shock intersection P so that the problem is one of steady flow. In these coordinates, the wall (or plane of symmetry) is moving downward at a constant velocity equal to the component of u_1 or u_4 tangential to S_M. A line PD can be drawn separating the fluid which has passed through S_{ine} and S_{ref} from that which has passed through S_M. The gas pressure (or liquid height) must be continuous across PD since it is not a shock. However, fluid which has passed through a single large shock has lost more flow energy than that which has passed through two smaller shocks with the same total change in pressure (or height); hence the velocity and density of a gas (or the velocity of a liquid) will be less below PD than above it. Thus, PD is a slip-stream. Schlieren photographs of gas flows have shown such a density discontinuity at PD.

Quantitative relations for such three-shock intersections can be obtained by applying the oblique shock equations to each of the shocks. The results for some ranges of interest have been tabulated by Polachek and Seeger both for compression shocks[12] and for hydraulic jumps.[13]

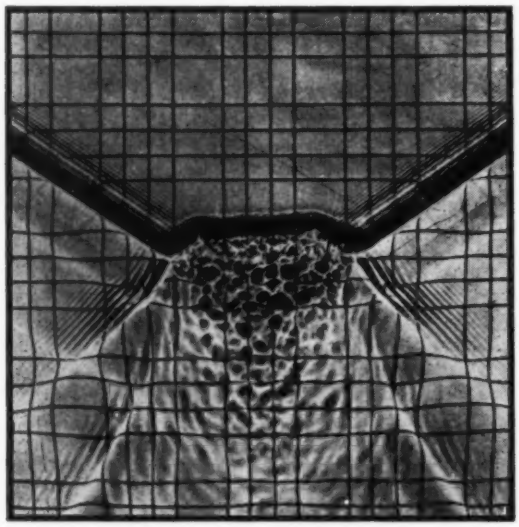

FIG. 9. Refraction photogram of medium-strength hydraulic-jump intersection ($\xi=0.45$, $\alpha=56°$).

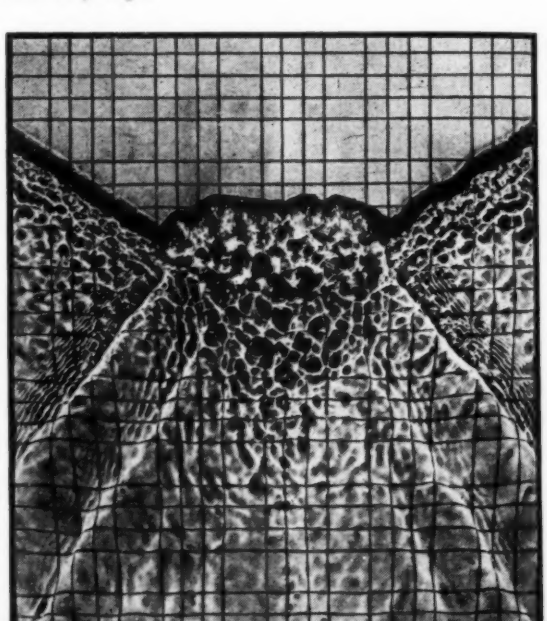

FIG. 10. Refraction photogram of strong hydraulic-jump intersection ($\xi=0.28$, $\alpha=56°$).

EXPERIMENTAL MEASUREMENTS

The purpose of the present experimental study was the investigation of regular and Mach intersections in hydraulic jumps. The apparatus used in the experiments was essentially that used by Einstein and Baird[14]

[12] H. Polachek and R. J. Seeger, Explosives Research Report No. 13, Navy Department, BuOrd, Washington, D.C. (February, 1944).

[13] H. Polachek and R. J. Seeger, Explosives Research Report No. 14, Navy Department, BuOrd, Washington, D.C. (August, 1944).

[14] H. A. Einstein and E. G. Baird, Progress Reports on the Analogy Between Surface Shock Waves on Liquids and Shocks in

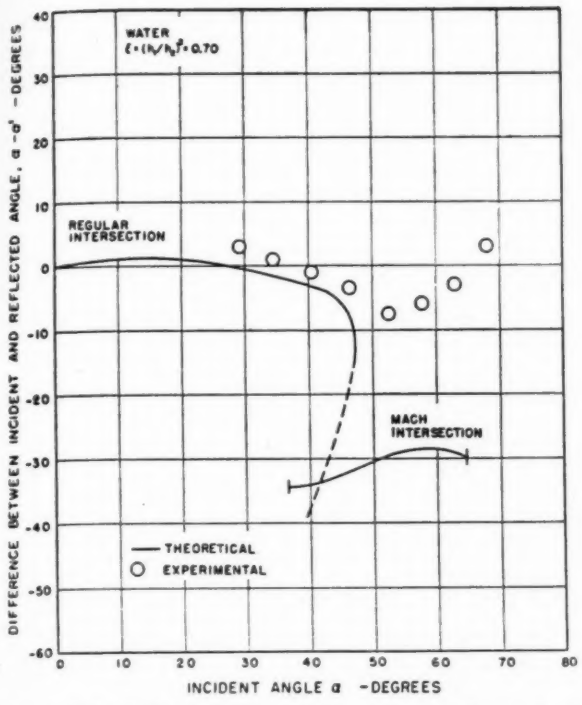

FIG. 11. Relationship between incidence and reflection angles for the intersection of weak hydraulic jumps ($\xi=0.70$).

in earlier investigations and is shown schematically in Fig. 6. The water table was approximately 5 ft. long, 4 ft. wide, and 2 in. deep. The liquid used in the tank was a 0.001 to 0.002 normal solution of manganous chloride in distilled water; the manganous chloride was added to the water so that the electrical conductivity of the solution would be more uniform. The hydraulic jumps were produced by two wave generators which were reservoirs into which a volume of water could be raised above the free surface; this volume of water could be discharged quickly through an adjustable horizontal slot 24 in. long located just below the free surface. Instantaneous water depths at any point on the water table could be determined by measuring the resistance between a pair of partially immersed platinum electrodes. A 1000-cycle alternating voltage was applied across the electrodes and the current which flowed was recorded by an oscillograph with an upper limit of about 100 cycles on its frequency response.

Typical wave depth records are shown in Fig. 7; the record is interrupted at intervals to give a zero reference line. The horizontal coordinate in these records is a time scale not a spatial scale, but a good estimate of the instantaneous wave form in space may be obtained by multiplying this time scale by the wave velocity.

In addition to a continuous record of the water depth at a point, an instantaneous refraction photogram of the entire field could be made. The bottom of the water tank was made of glass so that the light flash from a spark discharge across a $\frac{1}{8}$-in. gap could be transmitted upward through the water and the resulting refraction pattern could be recorded on photographic paper placed above the water surface.

RESULTS AND DISCUSSION

Measurements were first made to determine the shape and velocity of single hydraulic jumps produced with one wave generator alone; depths were determined with three pairs of electrodes. The initial water depth, h_1, ranged from about 0.5 cm for strong jumps to about 2 cm for weak jumps. Weak jumps were found to have an undulatory profile, stronger jumps had a smoother profile, and very strong jumps showed appreciable irregular roughness (cf. Fig. 7). These observations agree qualitatively with those of Bahkmeteff and Matzke[10] on hydraulic jumps larger by an order of magnitude than those reported here. The velocities of the hydraulic jumps were also measured and compared with the usual formula

$$u_1=\{g(h_1+h_2)h_2/(2h_1)\}^{\frac{1}{2}}, \qquad (2)$$

where h_1 is the water depth ahead of the jump and h_2 is the water depth behind the jump. The measured velocities of weak jumps agreed with the theoretical values within the experimental error of approximately two percent, but the measured velocities of the stronger jumps were up to five percent larger than the theoretical values. A possible explanation of this discrepancy is that the water released from the wave generator may tend to override the surface of the still water. The water

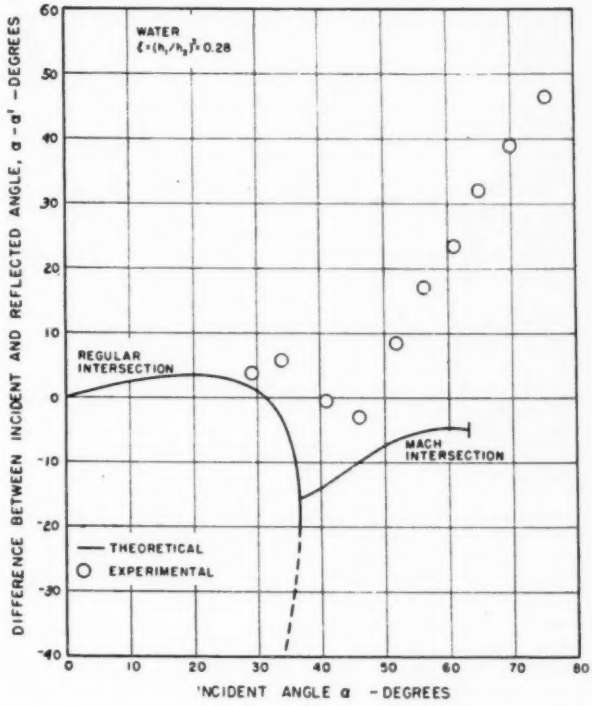

FIG. 12. Relationship between incidence and reflection angles for the intersection of strong hydraulic jumps ($\xi=0.28$).

Compressible Gases, Hydrodynamic Laboratories, California Institute of Technology (September 15, 1946 and July 30, 1947).

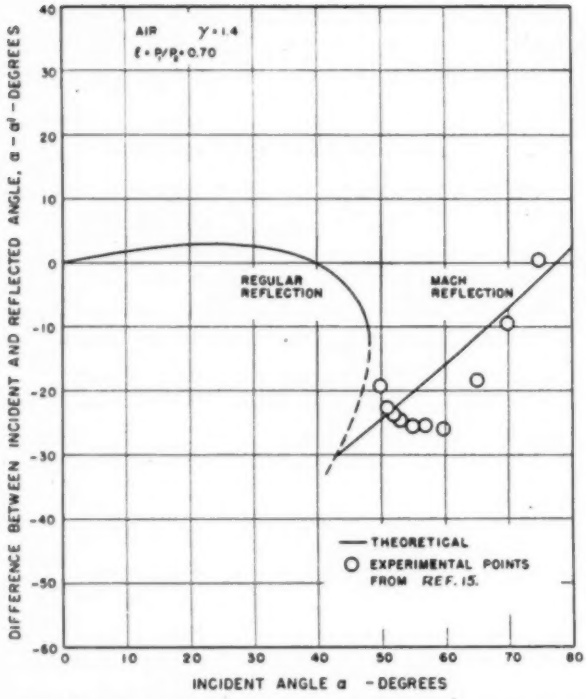

FIG. 13. Relationship between angles of incidence and reflection for the reflection of weak compression shocks ($\xi=0.70$) in air. The experimental points are from the data of Harrison and Bleakney (see reference 15).

in the jump would then have a higher velocity at the free surface than at the bed so that the simple theory would not apply.

A study was also made of the effect of the addition of surface-tension reducing agents to the water. When 0.1 percent by volume of lauryl isoquinolinium bromide was added to the water, the wave fronts formed became steeper, and the strong jumps became smoother. However, an appreciable amount of foam was produced which interfered with the photographic procedure. It was noted, moreover, that the amplitude of the undulations in the weaker jumps was increased. A weaker detergent solution, which consisted of 0.5 percent by volume of Kodak Photo-Flo (an aerosol solution), was found to give an intermediate effect: the undulations were enhanced only in the weakest waves, while the wave form of the strong waves was smoothed and steepened. The effect of these detergent solutions is shown in Fig. 8.

Measurements were made on the hydraulic jump intersections produced by two wave generators set at various angles. The jump strengths were determined by the resistance depth gauges while the general flow pattern was recorded on the refraction photograms (cf. Figs. 9 and 10).

Most of the measurements were made with jump strengths, $\xi\equiv(h_1/h_2)^2$, of 0.28, 0.45, and 0.70. For the medium and strong jumps ($\xi=0.45$ and 0.28), the aerosol solution was used, while for the weak jumps ($\xi=0.70$) detergent-free water was used, since this gave the best wave form. Figures 11 and 12 show the experimentally measured differences between the angle of incidence, α, and the angle of reflection, α', plotted against α for weak and strong jumps, respectively. Theoretical curves for the regular and Mach intersections are also shown. The agreement between theory and experiment is fair for the limited number of regular intersections investigated, but is definitely poor for the Mach intersections. There is little evidence of the sudden change in reflection angle which is predicted by the theory in passing from regular to Mach intersections.

These figures may be compared with Figs. 13 and 14, which give the corresponding data for intersections of compression shocks in air. The experimental points are taken from a report by Harrison and Bleakney.[15] The agreement between theory and experiment is seen to be better here, but an appreciable discrepancy remains, especially for the weaker shocks. There appears to be little resemblance between the trend of the disagreement in the liquid and gas cases.

One possible error in the use of the simple theory in the above comparisons is the neglect of the curvature of the Mach jump. The geometric effect of this curvature can be eliminated by considering only a small region about the triple jump intersection. Such a presentation of the data is shown in Figs. 15 and 16 where ω and ω' are the local angles which the incident and

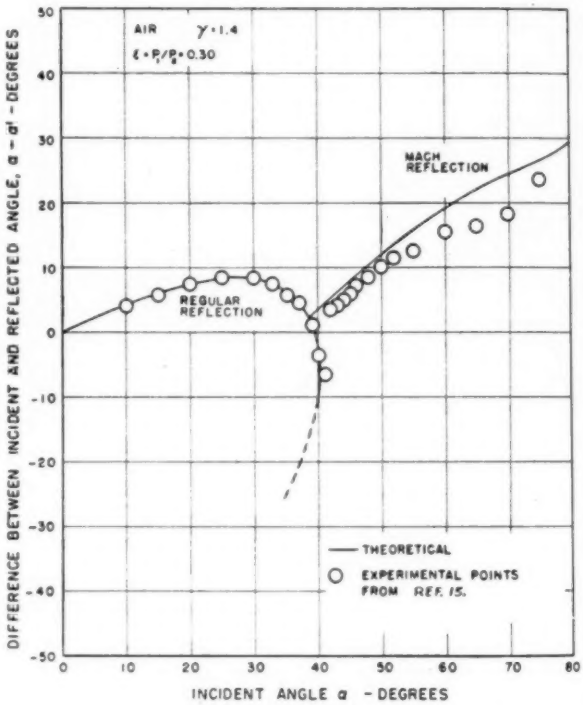

FIG. 14. Relationship between angles of incidence and reflection for the reflection of strong compression shocks ($\xi=0.30$) in air. The experimental points are from the data of Harrison and Bleakney (see reference 15).

[15] F. B. Harrison and W. Bleakney, "Remeasurement of reflection angles in regular and Mach reflection of shock waves," Physics Department, Princeton University (March, 1947).

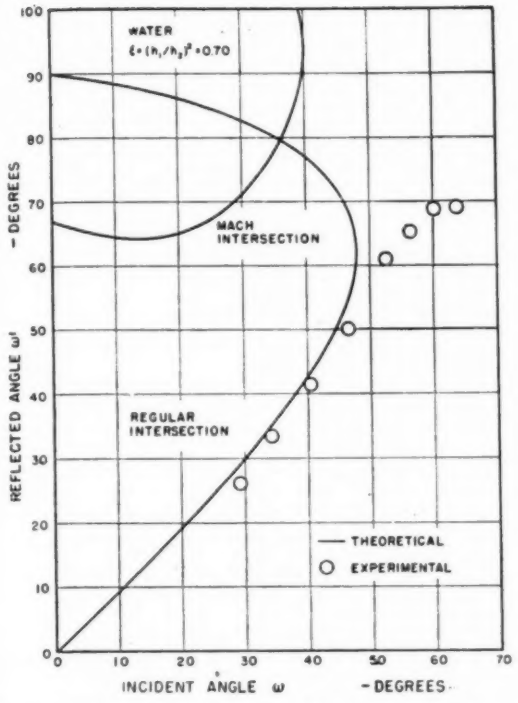

FIG. 15. Relationship between incidence and reflection angles, measured with respect to the direction of motion of the triple intersection, for the intersection of weak hydraulic jumps ($\xi=0.70$).

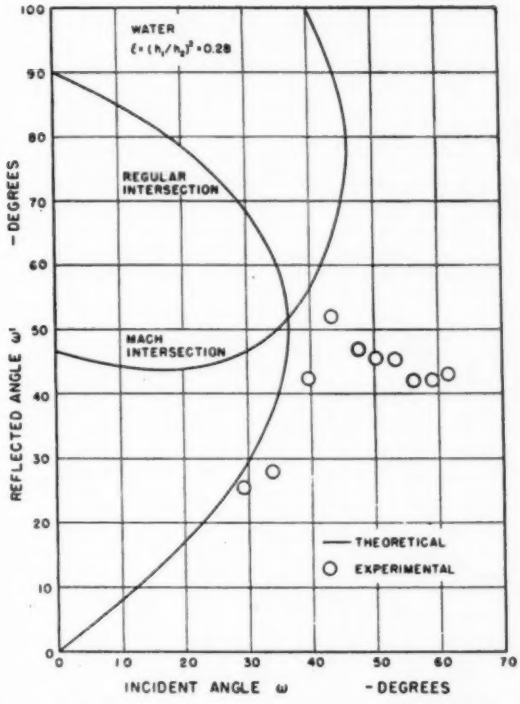

FIG. 16. Relationship between incidence and reflection angles, measured with respect to the direction of motion of the triple intersection, for the intersection of strong hydraulic jumps ($\xi=0.28$).

reflected jumps make with the direction of motion of the triple point. The weak-jump data shown in Fig. 15 show a disagreement with theory similar to that found by Harrison and Bleakney for air, although the discrepancy is somewhat greater for the hydraulic case. The strong-jump data shown in Fig. 16 show similar disagreement with theory, whereas Harrison and Bleakney found very little discrepancy in air.

Examination of the photograms indicates that the thickness of the Mach jump is not negligible compared with the radius of curvature of the Mach. The present theory does not take account of this possibility so that an error of unknown magnitude has been introduced. This difficulty does not arise in the aerodynamic case. The experimental observotions indicate an additional source of error in the simple theory: the liquid depth behind the reflected jump and the Mach jump is not uniform; in fact, there sometimes appears to be an additional hydraulic jump approximately parallel to, but behind, the reflected jump. Corresponding sources of error may exist in the aerodynamic case although direct observations on this possibility are not available.

One may conclude from the experimental observations discussed here that the deviations from the simple theory of hydraulic-jump intersections limit the use and validity of the hydraulic analog as a means for studying compression-shock intersections in gases.

An Optical Method for Measuring the Stress in Glass Bulbs

W. T. READ
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received October 6, 1949)

In the sealing and processing of electron tubes residual stresses are introduced into the glass parts. This paper formulates a practical optical method for locating and measuring the maximum tension in the walls of glass bulbs.

The basic principles of photoelasticity in three dimensions are applied to the case of a cylindrical shell. Theoretical results are verified experimentally for a tube in which the stress distribution is axially symmetrical.

Experimental results on tubes in which the stress distribution was unsymmetrical lead to the conclusion that the maximum tensile stress is proportional to the maximum retardation, the factor of proportionality being simply related to the form of the fringe pattern.

A quick and simple technique is described for making the necessary optical measurements.

The method presented herein has been used and found to be practical for routine laboratory testing of tubes in quantity.

I. INTRODUCTION

IN the sealing and processing of electron tubes, residual stresses are introduced in the bulb walls due to non-uniform heating and mismatch of thermal expansion coefficients. Failures resulting from these stresses constitute a serious problem especially in glass walled tubes. A quick and reasonably accurate method of determining the maximum stress in the bulb would permit the detection and rejection of overstressed tubes at the factory and aid in the improvement of sealing techniques.

The usual procedure is to examine the tube in a polariscope and judge the stress magnitude by observation of relative phase retardation. Because of the three-dimensional state of stress in the tube wall the quantitative conversion of the observed retardation to stress magnitude involves considerations beyond those commonly employed. It was the purpose of this investigation to discover the relation between the observed retardation and the maximum stress in the cylindrical portion of the bulb and to develop a simple laboratory procedure for making the necessary measurements.

We begin with a brief review of the stress optical relations in the general three-dimensional case.

Upon entering a doubly refracting material such as stressed glass, the incident light is resolved into two waves, plane polarized at right angles to one another and traveling with different velocities. In a material as weakly birefringent as glass the two waves may be assumed to travel in the same direction and to have the same wave normal. In Fig. 1 let the wave normal be parallel to the y axis so that the wave front is parallel to the xz plane, f being the direction of polarization of the faster moving wave and s, of the slower moving wave. We designate φ as the (variable) angle between f and the x axis. The angle φ and the relative velocities of the two waves are determined by the directions and magnitudes of the secondary principal stresses in the plane of the wave front. The axes of secondary principal stress are the normals to the two perpendicular planes through the y axis on which the shearing stress vanishes, and the normal stress is stationary. The secondary principal stresses, p and q, are respectively the maximum and minimum values of the normal stress in the xz plane.

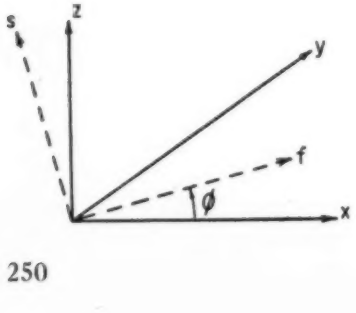

FIG. 1. Directions of polarization in plane of wave front.

We shall use the following two stress-optical laws which apply to weakly birefringent materials such as glass: (1) at each point in the stressed material the polarizing axes f and s in the plane of the wave front are along the axes of secondary principal stress, and (2) the difference between the reciprocals of the velocities of the two waves is proportional to the difference between the secondary principal stresses, i.e., to $p-q$. For most materials the direction of the maximum normal stress p is along the f axis.

The angle φ (Fig. 1) is, therefore, the angle between the direction of maximum normal stress and the x axis. Letting σ_x and σ_z be the normal components of stress in the x and z directions and τ_{xz} be the shearing component we have the relations[1]

$$\tan 2\varphi = (2\tau_{xz})/(\sigma_x - \sigma_z) \tag{1}$$

$$p-q = [(\sigma_x-\sigma_z)^2 + 4\tau_{xz}^2]^{\frac{1}{2}}. \tag{2}$$

If the light is monochromatic the vibration at any point in the glass is defined by the difference in phase ψ between the components along the f and s axes and by the ratio of amplitudes, $\tan\omega$. The two functions of y, ω and ψ, are given in terms of the functions φ and

[1] R. D. Mindlin, J. App. Phys. **10**, No. 4, 229 (April, 1939).

$p-q$ by the following two simultaneous first-order differential equations with variable coefficients:[2]

$$(d\omega/dy)+(d\varphi/dy)\cos\psi=0$$

$$\frac{\tan 2\omega}{2}\left[\frac{d\psi}{dy}+\frac{2\pi C}{\lambda}(p-q)\right]-\frac{d\varphi}{dy}\sin\psi=0, \qquad (3)$$

where λ is the wave-length of the incident light and C is the relative stress optical coefficient.

These equations were derived by F. Neumann from simple kinematical considerations involving approximations which are valid for photoelastic studies.[3] In the standard two-dimensional case the stress distribution is nearly constant through the thickness of the model so that rotation of the polarizing axes along the path of the ray may be neglected and Neumann's equations may be easily integrated to give $\omega=$ const and the standard two dimensional formula

$$\psi/2\pi=C/\lambda(p-q)d,$$

where d is the thickness of the model. In the general three-dimensional case there is rotation of the secondary principal axes and the exact solution is impractical. However, in some cases the equations can be considerably simplified subject to certain limitations on the quantities involved. Denoting by $R/2$ the ratio of rate of rotation of polarizing axes to rate of relative phase retardation referred to stationary axes and substituting into Neumann's first equation gives

$$d\psi/dy=-(2\pi C)/\lambda(p-q)[1+R\cot 2\omega\sin\psi]. \qquad (4)$$

If the rate of rotation of the secondary principal axes is small compared to the rate of relative phase retardation and if the amplitudes of the two components are equal on entry and remain approximately equal (so that ω remains close to $\pi/4$ radians) then $R\cot 2\omega$ is negligibly small compared to unity and we may use the simple formula

$$d\psi/dy=-(2\pi C)/\lambda(p-q). \qquad (5)$$

However here $p-q$ is in general an unknown function of y so that we shall leave the equation in differential form. Thus subject to the conditions on R and ω the rate of relative phase retardation depends only on the secondary principal stress difference, and not on the rotation of the polarizing axes. The exact form of the equations and the effect of large $R\cot 2\omega$ are discussed for simple states of stress in Appendix I.

II. PHOTOELASTIC THEORY OF SHELLS

In this section we shall regard the bulb as a cylindrical shell and use the concepts and approximations of shell theory. Figure 2 shows a cross section of the cylinder perpendicular to the axis and a typical element of volume cut out of the tube wall. The tube is in a field of parallel rays and is surrounded by a liquid medium having the same index of refraction as the glass of the tube wall so that the light rays are not refracted on entry and exit. We designate a as the radius of the middle surface of the shell and h as the half-thickness. As before we take the y axis parallel to the light rays so that the xz plane is parallel to the wave front. We shall also use a cylindrical coordinate system constructed in the usual way with respect to the cylinder, except that the radial coordinate is measured from the middle surface of the shell.

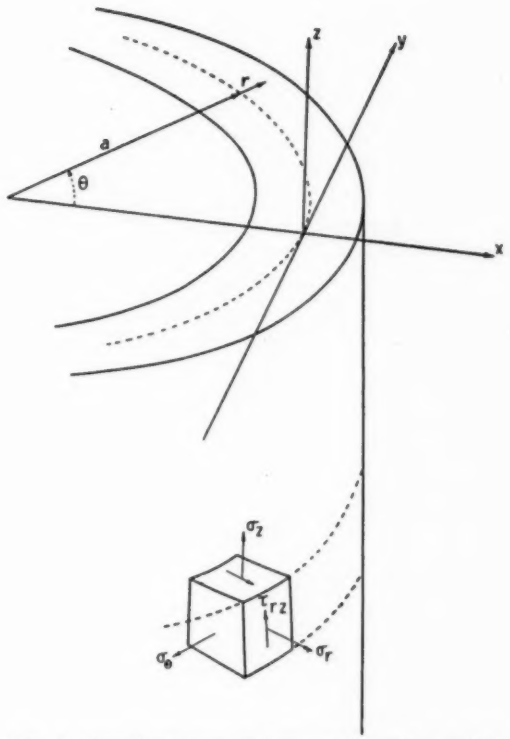

FIG. 2. Coordinate axes and stresses in wall of cylinder. Light rays are parallel to y axis.

The lower part of Fig. 3 shows the stress components in cylindrical coordinates. In shell theory the following assumptions are made regarding the stresses: the radial stress, σ_r, is identically zero, the circumferential stress, σ_θ, is constant, the axial stress, σ_z, is a linear function of r which vanishes on the middle surface, and the shearing stress τ_{rz} is a parabolic function vanishing on the inner and outer surfaces as required by the boundary conditions. The relations between the stresses σ_θ and τ_{rz} and the stresses σ_x and τ_{xz} in the plane of the wave front are

$$\tau_{xz}=\tau_{rz}\cos\theta$$
$$\sigma_x=\sigma_\theta\sin^2\theta.$$

Substituting into the formula for φ (Eq. (1))

$$\tan 2\varphi=\frac{2\tau_{rz}\cos\theta}{\sigma_\theta\sin^2\theta-\sigma_z}. \qquad (6)$$

[2] E. G. Coker and L. N. G. Filon, *Photo-Elasticity* (Cambridge University Press, London, 1931), p. 256.

[3] For the relation between Neumann's equations and Maxwell's electromagnetic equations see R. D. Mindlin and L. E. Goodman, "The optical equations of three-dimensional photoelasticity," J. App. Phys. **20**, No. 1, 89–94 (January, 1949).

Since τ_{rz} must be zero on the inner and outer surfaces the angle φ between the polarizing axis of the faster moving wave and the x axis will be zero on exit and entry. Thus at the beginning and end of the path of a ray through the wall of a tube in an axially symmetrical state of stress the two polarizing axes are respectively along and perpendicular to the axis of the cylinder.

Inside the tube wall τ_{rz} will in general be non-vanishing and there will be some rotation of the secondary principal axes along the path of the ray. However, in the photoelastic study of stresses in electron tube bulbs we are interested only in the stresses at the point where failure is most likely to occur. Hence we wish to investigate the stress distribution where σ_z is a maximum and in particular to measure the maximum. Since glass is weak in tension, breakage will result from either excessive positive axial or circumferential stress. In the tubes considered here the axial stress was the critical

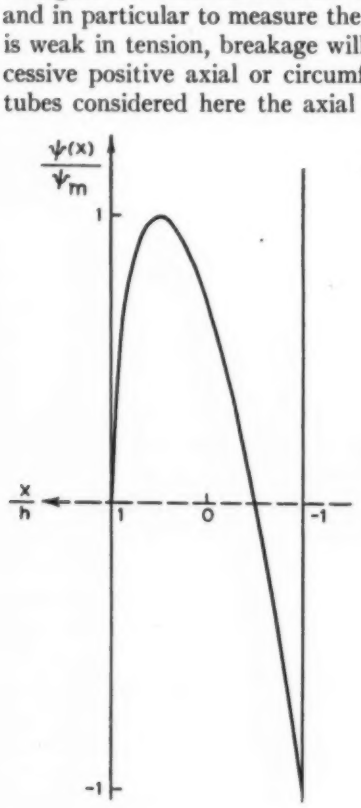

FIG. 3. Theoretical fringe pattern in bulb wall for symmetrical stress. Fringe displacement is proportional to phase difference.

one, as indicated by breakage results and by independent measurements of circumferential stress. The method of measuring σ_θ is described in Appendix II.

The axial stress is given by

$$\sigma_z=\sigma(r/h),$$

where σ is a function of z and is the maximum of σ_z for a given cross section. The absolute maximum of σ_z occurs where σ has its maximum value, and therefore where $d\sigma/dz=0$. Using this condition and the equation of equilibrium in the axial direction,[4]

$$\frac{\partial\sigma_z}{\partial z}+\frac{1}{r+a}\frac{\partial}{\partial r}(r+a)\tau_{rz}=0 \tag{7}$$

[4] S. Timoshenko, *Theory of Elasticity* (McGraw-Hill Book Company, Inc., New York), p. 309. (Timoshenko's r is our $r+a$.)

we have $(r+a)\tau_{rz}=$constant. However, from the boundary conditions, τ_{rz} is zero at the inner and outer surfaces, so that $\tau_{rz}=0$ throughout the thickness of the wall, and from (6) $\varphi=0$ or $\pi/2$ over the cross section. We have now established the following useful result: in a cylindrical shell there is no rotation of the polarizing axes along the path of a ray lying in the cross section where the axial stress reaches its maximum value. Therefore Eq. (5) applies exactly and ω is constant and equal to the angle between the x axis and the plane of polarization of the incident light.

To discover the meaning of the observed retardation in terms of stress we use (5). In the cross section of maximum axial stress $\tau_{rz}=0$ so that (2) simplifies to

$$p-q=\sigma_x-\sigma_z=\sigma_\theta\sin^2\theta-\sigma_z.$$

Letting σ_m be the maximum value of σ_z we have

$$d\psi/dy=2\pi C/\lambda[(\sigma_m/h)r-\sigma_\theta\sin^2\theta]. \tag{8}$$

To integrate this equation we express r and θ in terms of the variable y and the parameter x which is constant for a given ray and is equal to the minimum distance between the ray and the middle surface. When a is large compared to r and x we have the approximations:

$$y^2/2a=r-x \quad \sin\theta=y/a.$$

As the ray passes through the tube wall y varies between $-[2a(h-x)]^{\frac{1}{2}}$ and $+[2a(h-x)]^{\frac{1}{2}}$. Since the incident light is monochromatic and plane polarized there is a single λ and the phase difference on entry is zero. The phase difference on exit $\psi(x)$ is a function of the parameter x given by the integral of Eq. (8)

$$\frac{\psi(x)}{2\pi}=\frac{CL\sigma_m}{\sqrt{2}\lambda}\left[\frac{x}{h}+\frac{1}{3}\left(1-2\frac{\sigma_\theta}{\sigma_m}\frac{h}{a}\right)\left(1-\frac{x}{h}\right)\right][1-(x/h)]^{\frac{1}{2}},$$

where $L=2(4ah)^{\frac{1}{2}}=2[(\text{diameter})\times(\text{wall thickness})]^{\frac{1}{2}}$ is the length of the path of a ray tangent to the inner surface of the bulb.

In all of the tubes observed σ_θ was of the same order of magnitude as or less than σ_m, and in electron tubes $h\ll a$ so that we can neglect $2(\sigma_\theta/\sigma_m)(h/a)$ in comparison with unity. Then

$$\frac{\psi(x)}{2\pi}=\frac{CL\sigma_m}{3\sqrt{2}\lambda}\left(1+2\frac{x}{h}\right)[1-(x/h)]^{\frac{1}{2}}. \tag{9}$$

We now have another useful result concerning cylindrical shells; namely, that in the cross section where the axial stress is a maximum the relative phase retardation depends on only the axial stress provided that the maximum axial stress is of the same order of magnitude as or greater than the circumferential stress. If σ_θ is much greater than σ_z the technique described in Appendix II should be used to measure σ_θ and the formula for σ_m connected accordingly.

Figure 3 is a plot of $[\psi(x)]/\psi_m$ as given by Eq. (9), where ψ_m is the maximum of $\psi(x)$.

Since we have assumed an axially symmetrical distribution of stress the first experiments to check the theory were made on a type of tube in which the fringe pattern remained constant as the tube was rotated about its axis. The phase difference in the cross section of maximum axial stress was measured using polarized light and a quartz wedge. Figure 4 (center fringe) shows the fringe pattern and tube wall focused in the same plane. The retardation caused by passage through the tube wall is proportional to the displacement of the fringe in the vertical direction, the distance between fringes corresponding to a retardation of 2π radians or one wave-length. Since the stress does not vary appreciably in the vertical direction in a distance equal to the maximum fringe displacement the fringe pattern itself is approximately a plot of the measured $\psi(x)$ *vs.* x.

Comparison of Figs. 3 and 4 shows that the theoretical curve of Fig. 3 derived from shell theory agrees very well with the pattern actually observed.

The formula for $\psi(x)$ Eq. (9) may be used to relate the measured maximum phase difference ψ_m to the maximum axial stress σ_m. Letting $\sigma_0=\lambda/CL$ be the stress which would produce a relative retardation of one wave-length, that is a phase difference of 2π radians, in the distance, L, traveled by the innermost ray of the tube wall, and letting $R_m=\psi_m/2\pi$ be the relative retardation in wave-lengths we have

$$\sigma_m=3\sigma_0R_m. \tag{10}$$

If λ is expressed in angström units, L in millimeters and C in 10^{-2} brewsters than σ_0 and σ_m will be in kilograms per square millimeter.

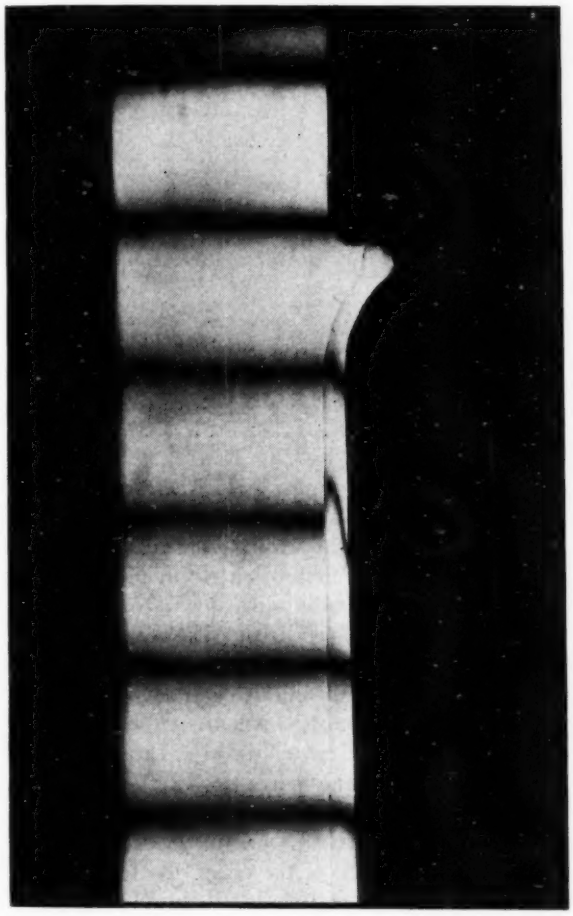

FIG. 4. Actual fringe pattern in bulb wall for symmetrical stress.

III. ROTATION OF POLARIZING AXES—UNSYMMETRICAL STRESSES

We shall now consider in more detail the formation of the fringe pattern in Fig. 4. After emergence from the tube the originally plane polarized light is elliptically polarized. Since the f and s axes on exit are parallel to the z and x axes which in these experiments are vertical and horizontal respectively, the quartz wedge is oriented with its axes also vertical and horizontal. The relative retardation produced by the wedge varies linearly in the vertical direction so that for any value of x there is always a point on the wedge where the relative retardation caused by the tube $\psi(x)$ is equal and opposite to that caused by the wedge. At this point the light reaching the analyzer will be plane polarized at an angle $\omega(x)$ to the x axis. If the analyzer is set at an angle of $\omega(x)+(\pi/2)$ radians the light will be extinguished and the point will appear as a dark spot in the observed fringe pattern, the displacement of the point being proportional to $\psi(x)$. However this condition cannot be simultaneously satisfied for all x unless $\omega(x)$ is constant. For a variable $\omega(x)$ we would observe not a continuous fringe in the tube wall but an isolated dark spot which would move across the thickness of the tube wall as the analyzer was rotated.

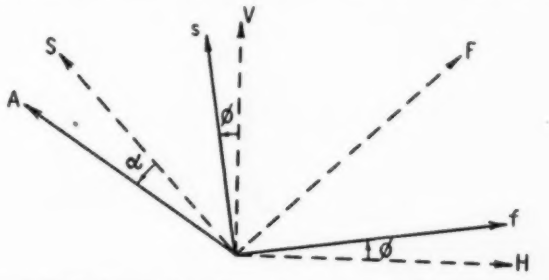

FIG. 5. Optical axes of specimen, quarter wave plate and analyzer.

The theory of cylindrical shells predicts that where the axial stress is a maximum $\omega(x)$ will be constant and there we would expect to observe a continuous fringe. Actually, however, continuous fringes could be observed at every cross section indicating that throughout the bulb the deviation of ω from the initial value of $\pi/4$ radians is small enough to remain within the band largely extinguished by the analyzer. Rotating the analyzer obscured rather than improved the fringe pattern as a whole. Also the intensity at any point on the fringe was increased by rotating the analyzer by more than $\pm5°$, indicating that the variation in ω is

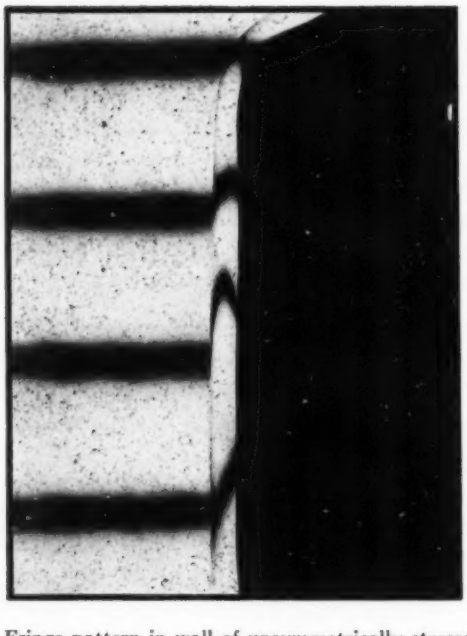

FIG. 6. Fringe pattern in wall of unsymmetrically stressed bulb. Maximum fringe displacement indicates maximum stress.

within these limits. From Eq. (3) the small variation in ω implies that the rotation of the polarizing axes is small which in turn implies that the shearing stress τ_{rz} is small compared with the axial stress σ_z. Thus the results obtained for the cross section of maximum axial stress apply approximately to every cross section and the phase difference at every level may be regarded as a function of the axial stress.

Where the stress distribution is not axially symmetrical the problem is considerably more complicated and we have the stress components $\tau_{r\theta}$ and $\tau_{\theta z}$ in addition to those shown in Fig. 2. However, even in a type of tube in which the stress distribution arose from local thermal effects and varied rapidly and erratically in both the axial and circumferential directions, continuous fringes could be observed at every level with the quartz wedge and cross polaroids. Again rotating the analyzer obscured the fringes. Thus even for this extreme case it was possible to neglect the effects of the shearing stresses and regard the observed retardation as a function of axial stress.

In the general case the maximum tension σ_m is related to the maximum relative phase retardation R_m by $\sigma_m = W\sigma_0 R_m$, where σ_0 is defined as in the preceding section and W is a dimensionless weighting factor depending on the form of the fringe pattern, that is on the variation of ψ on exit with x. As we have seen (Eq. (10)) the stress distribution assumed in shell theory gives $W=3$. The minimum value of W is unity and occurs when the axial stress is constant over the cross section. If we make the approximation that over the path of a ray in the bulb wall the stress depends only on r we can calculate the fringe patterns corresponding to different simple variations of σ_z with r. Comparing the calculated patterns with those observed we determine the range of variation of W and its dependence on obvious features of the fringe pattern. In the tubes tested at the Bell Laboratories W was found to vary from 1 to 3 and to depend mostly on the position of the observed maximum phase difference between the outer and inner surfaces of the bulb wall. For routine testing W was taken as 3 when the maximum was observed in the outer part of the bulb wall, as in Fig. 4, and as 2 for a maximum appearing in the center or inner half of the wall, as in Figs. 6 and 7. If the fringe displacement had no maximum but increased monatonically from the outer to inner surface of the wall as in Fig. 6 (bottom fringe) then W could safely be taken as unity.

Although the quartz wedge gives the most information for experimental studies, we shall see in the following section that the most practical routine method of identifying overstressed bulbs involves the use of a quarter wave plate or de Sénarment compensator. The quarter wave plate is first oriented to give minimum intensity with crossed polaroids at 45° to the vertical. The tube is then inserted with axis vertical between the polarizer and quarter wave plate and the analyzer rotated to give minimum intensity at a specified point. If $\varphi=0$ and $\omega=\pi/4$ on exit from the tube the rotation of the analyzer in radians will be equal to one-half the phase difference for the corresponding ray. By this method it is possible to obtain measurements of phase difference accurate to a few degrees. Since the deviations in ω and φ are also of the order of a few degrees it is interesting to ask whether the deviations in ω and φ are augmented or diminished in their effect on the accuracy of the measurement of phase difference.

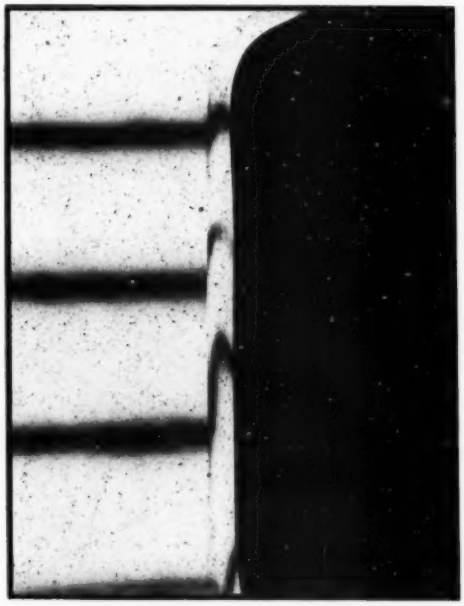

FIG. 7. Same as Fig. 6 except quartz wedge is raised to show fringes at different levels.

First we shall consider the case where $\varphi=0$ on exit and investigate the effect of taking $\omega\neq\pi/4$.

In Fig. 5 F and S are the fast and slow polarizing axes of the quarter wave plate which for $\varphi=0$ are at 45° to the f and s axes on exit. If α is the angle between the axis of the analyzer A and the S axis then it follows from simple trigonometry that the intensity of the light transmitted by the analyzer is proportional to

$$1-\cos 2[\omega-(\pi/4)]\cos(2\alpha-\psi).$$

If $\omega=\pi/4$, the light is extinguished when $2\alpha=\psi$. If $\omega\neq\pi/4$ it is impossible to obtain extinction. However, we do not set for absolute extinction but for minimum intensity which for low intensities is indistinguishable by eye from absolute extinction. It is evident from the above formula that the intensity of the light observed through the analyzer will always be a minimum when $\alpha=\frac{1}{2}\psi$ independently of the value of ω. We thus have the result that if the quarter wave plate is oriented

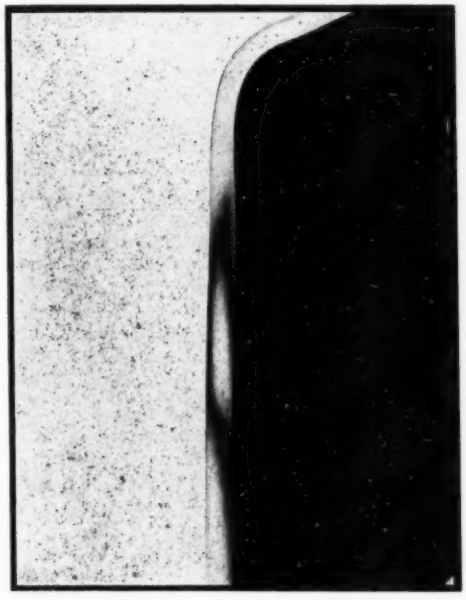

FIG. 8. Fringe pattern with quarter wave plate. Ring in bulb wall encloses region of critical stress.

with its axes at 45° to the polarizing axes of the tube on exit then the formula $\psi=2\alpha$ will be correct independently of the rotation of the polarizing axes along the path of the ray inside the tube wall.

Now consider the case where φ is not zero on exit. If we denote by ψ' the difference in phase between the H and V components we have from the preceding result that $\psi'=2\alpha$ for minimum intensity. That is, the de Sénarment compensator correctly measures the difference in phase between the H and V components. The error in the measurement of ψ is therefore $\psi'-\psi=2\alpha-\psi$. After considerable trigonometric manipulation we derive for the first-order term in $\psi'-\psi$

$$\psi'-\psi=2\sin\psi\sin\varphi\frac{\sin(\varphi+2\omega-\pi/2)}{\cos 2(\varphi+\omega-\pi/4)}$$

or neglecting second-order terms in φ and $\omega-\pi/4$ in comparison with unity we have for the upper limit of the absolute error

$$|\psi'-\psi|\leqq|2\varphi^2+4\varphi(\omega-\pi/4)|.$$

Thus the error in the measurement of relative phase retardation is of second order in the deviations in ω and φ on exit and may therefore be considered negligible.

IV. TECHNIQUE OF MEASUREMENT

The de Sénarment compensator or quarter wave plate is free from a serious limitation of the quartz wedge, which gives fringes at only three or four levels in the tube wall. With the rapid and erratic variation of stress in the axial direction there is a high probability that maxima may fall between the fringes and consequently not be recorded. The de Sénarment compensator not only covers the whole extent of the tube in the axial direction but also provides the basis for a simple and practical technique for routine testing.

We have seen that when the analyzer is set at an angle α (to the S axis of the quarter wave plate—Fig. 5) a dark spot is observed in the resulting fringe pattern wherever the phase difference is 2α.

Consider the case where the analyzer is set so that 2α is equal to the maximum phase difference in the whole bulb. Then a single dark spot or fringe will be observed at the maximum and the rest of the bulb wall will

FIG. 9. Shrinking of ring about critically stressed region as analyzer is rotated.

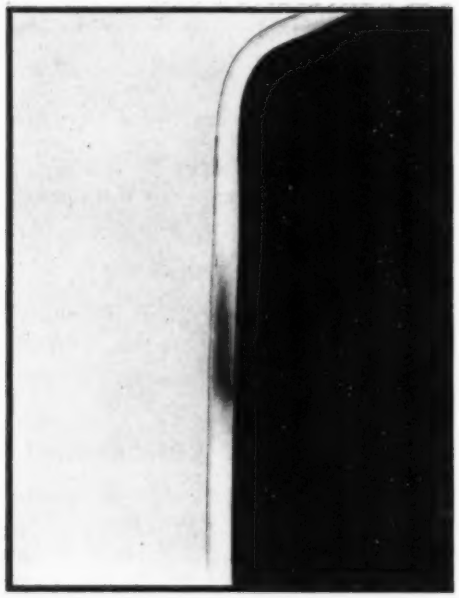

FIG. 10. Ring shrunk to single dark spot at level of maximum stress. Rotation of analyzer gives magnitude of maximum stress.

appear light as in Fig. 10. When the analyzer is set so that 2α is slightly less than the maximum phase difference the maximum will appear as a light spot enclosed by a closed dark fringe or ring on which the phase difference is 2α, Figs. 8 and 9. By rotating the analyzer so that α increases, the ring is caused to shrink to a single dark spot. The setting of the analyzer then gives the maximum phase difference.

Figures 6 and 7 are taken with a quartz wedge and illustrate a fringe pattern of typical form but somewhat larger than average amplitude. The maximum phase difference is seen to be approximately $\frac{2}{3}$ of a wavelength or $4\pi/3$ radians. This corresponds to a setting of $\alpha=120°$ for the de Sénarment compensator.

Figures 8 to 10 are taken with the de Sénarment compensator with $\alpha=60°$ in Fig. 8 $\alpha=85°$ in Fig. 9 and ending with the final setting of 120° in Fig. 10, which gives a single dark spot indicating a maximum phase retardation of 240° or $4\pi/3$ radians in agreement with the quartz wedge. The progressive shrinking of the dark ring can be clearly seen as the analyzer is rotated to approach the setting corresponding to the maximum retardation.

We now have established the following method for locating overstressed regions in the walls of glass bulbs with a de Sénarment compensator. When a limit on allowable stress has been set, the corresponding limit on phase difference is calculated (from σ_0) for the outer part of the tube wall where the weighting factor is 3. With the analyzer set for this phase difference the bulb is rotated once around its axis and the fringe pattern observed. If a dark ring appears in the outer part of the wall then the tube is rejected since the ring must enclose a maximum phase difference corresponding to a tensile stress above the allowable limit. If a dark ring appears in the center or the inside of the center of the tube wall, where the weighting factor is 2, then the analyzer setting is increased by a factor of $\frac{3}{2}$. If this causes the dark ring to shrink to a point and disappear then the tube is safe, but if a light spot still remains in the shrunken ring then the tube is rejected.

This procedure has been found to be practical for routine testing of tubes in quantity and is now in use at the Bell Telephone Laboratories at Allentown, Pennsylvania.

The experimental work involved in this investigation was carried out jointly by the author and Mr. J. E. Clark of the Electronic Apparatus Development Department of the Bell Laboratories. The photographs shown in Figs. 6 to 10 were made by R. D. Mindlin.

APPENDIX I

We shall now consider the case where both $d\varphi/dy$ and $p-q$ in Eq. (3) are constant. The exact solution for this case has been obtained by D. C. Drucker and R. D. Mindlin[5] using a dynamical theory of light.

We wish to study the dependence of the relative phase retardation ψ on the magnitude of R and the direction of polarization of the incident light. For this purpose we shall use Eqs. (12)

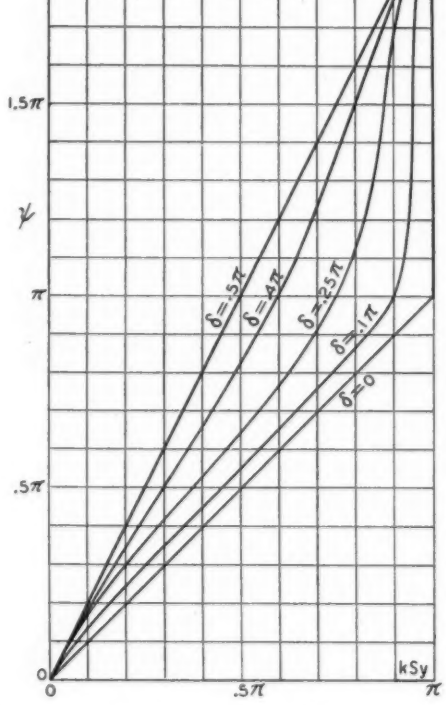

FIG. 11. Variation of phase difference with distance for several simple cases of constant principal stress difference.

[5] D. C. Drucker and R. D. Mindlin, "Stress analysis by three-dimensional photoelastic methods," J. App. Phys. 11, 724 (November, 1940).

of Drucker and Mindlin. These equations were obtained by Drucker and Mindlin upon inserting the assumption of weak birefringence in their exact solution. It can be shown that the identical formulas can be obtained by a direct solution of Neumann's equations. From these equations the difference in phase of the two waves (our ψ) is related to the distance along the path of the ray (our y) by

$$\tan\psi = \frac{\tan 2kSy}{S}\left[\frac{\tan\delta + \tan kSy}{\tan\delta + \tan 2kSy}\right], \tag{11}$$

where

$$k = \pi C/\lambda(p-q)$$
$$S = (1+R^2)^{\frac{1}{2}}$$
$$\tan\delta = \frac{S}{R\cot 2\omega_0}$$

and ω_0 is the angle between the direction of polarization of the incident light and the f axis on entry. Here we shall consider that $R^2 \ll 1$ so that we may take $S=1$.

When the incident light is plane polarized at 45° to the f axis on entry, $R\cot 2\omega_0 = 0$ as in the simple two-dimensional case (Eq. (3)). However, as the direction of polarization of the incident light approaches the direction of one of the secondary principal axes on entry, then $\cot 2\omega_0$ becomes infinitely large, and since there is always some rotation of the secondary axes, δ approaches 0. In this case the slope of the ψ *vs.* kSy curve approaches infinity in the range $\pi \leqq \psi \leqq 2\pi$ while in the range $0 \leqq \psi \leqq \pi$ the average slope is just one-half the average slope in the previous case. Figure 11 shows plots of ψ *vs.* kSy for several values of the parameter δ. In each case we take $S=1$.

Thus in this simple example where $p-q$ and R are constant and $R^2 \ll 1$ the simple two-dimensional formula gives accurate results for the relative phase retardation when the incident light is polarized at 45° to the secondary principal axes on entry. However, when the direction of polarization of the incident light coincides with one of the secondary principal axes then the simple two-dimensional formula gives values of ψ which may be in error by a factor of 2.

If $S \neq 1$ then the form of the curves in Fig. 11 will be altered somewhat but the curves will still cut the lines $\psi=0$, $\psi=\pi/2$, $\psi=\pi$, $\psi=3\pi/2$, and $\psi=2\pi$ at the same values of kSy.

It should be noted that the points of equal phase will occur at intervals of π/kS independently of the direction of polarization of the incident light. If $p-q$ is calculated from measurements of the distances between points where the light is plane polarized (as in Weller's method) then the simple two-dimensional formula is correct within a factor of S (see reference 5, p. 729).

APPENDIX II

Measurement of Circumferential Stress

In this section we shall describe the method of measuring σ_θ. This method was suggested by R. D. Mindlin.

For this measurement the tube was oriented with its axis vertical, as before, but was not immersed in the liquid. Fig. 12 shows part of the tube wall bounded by sections of the outer surface A and the inner surface B. The two sets of reflected rays have approximately the same intensity since each represents the result of one reflection, and little energy is lost on refraction.

From the laws of geometrical optics it follows that the two

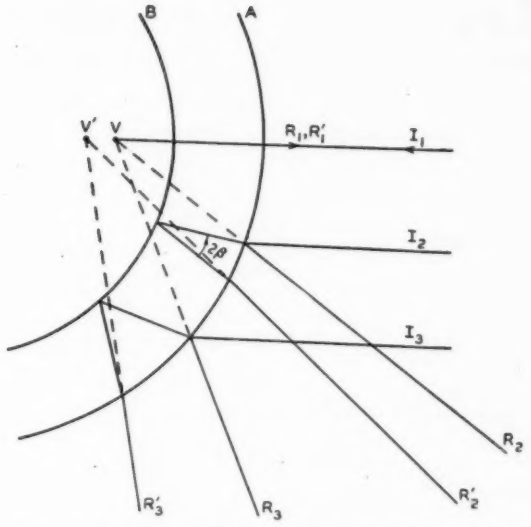

FIG. 12. Technique of measuring circumferential stress.

virtual sources V and V' are separated by a distance of $1/2n$ times the thickness of the wall where n is the index of refraction, that is the inner and outer surfaces of the tube are effectively cylindrical mirrors and the two virtual sources appear as lines parallel to the axis of the tube.

The light reflected from the outer surface of the tube remains plane polarized. However, the light reflected from the inner surface of the tube has twice passed through the tube wall and has been resolved into components which emerge out of phase.

The angle of incidence and index of refraction determine the angle β between the rays traveling inside the tube wall and the radial direction. The secondary principal stress difference in the plane of the wave front is therefore $\sigma_\theta \cos^2\beta - \sigma_z$. This gives for the phase difference ψ

$$\psi/2\pi = 2\int_{-h}^{h} \frac{C}{\lambda}(\sigma_\theta \cos^2\beta - \sigma_z)\frac{dr}{\cos\beta}.$$

Since σ_z is an odd function of r it disappears in the integration and we have for the constant circumferential stress σ_θ

$$\sigma_\theta = \frac{\lambda}{4hC\cos\beta}\frac{\psi}{2\pi},$$

where as before h is the half-thickness of the tube wall. The phase difference can be measured with a quarter wave plate and analyzer inserted in the path of the reflected rays. This technique may be used for routine testing of tubes in quantity since it is only necessary to set the analyzer for the retardation corresponding to the critical circumferential stress, rotate the tube once around its axis and observe whether the virtual line source is extinguished at any point.

However, this technique cannot be used to measure the circumferential stress in any tube where the average axial stress over the thickness is not zero. However, by this technique it was shown that there was no rotation of the polarizing axes in the plane of the wave front thus indicating that the shearing stress $\tau_{\theta z}$ is negligible.

Letters to the Editor

Note on the Frictionless Bearing for Small Angular Deflections

J. A. HARINGX
Philips Research Laboratories, Eindhoven, Netherlands
October 21, 1949

IN the paper by Helmut Schlitt,[1] I found an interesting description of a new type of bearing for small angular deflections consisting of two concentric rings connected to each other by three radial flat springs. By choosing the radii of the rings to be equal and omitting one of the springs we obtain a construction that is known as a cross-spring pivot. In a recent paper,[2] I published the results of a calculation in respect to this constructional element and emphasized the remarkable phenomenon of its maximum rigidity for definite values of the compressive forces exerted upon the flat springs. In the paper by Schlitt I missed the occurrence of such a maximum.

In Figs. 3 and 4 of the paper[1] the curves near the vertical coordinate axis both point upward. This result is to be doubted for the following reasons. Although the separate treatment of the flat springs for tensile and compressive loads is necessary owing to the different mathematical expressions used (the hyperbolic functions change into the cyclic functions), the transition from the tensile to the compressive force is actually a gradual one, as will be evident from a physical point of view. Thus, when drawn in one single graph where the two opposite meanings of the axial load P_s are taken into account, the curves of the two separate figures for tensile and compressive loads must gradually pass into each other. Apart from the difference in the ordinates for $P_s=0$ caused by the differently chosen values of R_1/R_2, the curves may not show the sharp peak arising here.

For the case that R_1 and R_2 are equal the actual relation between the bending moment about the center and the axial force may be taken from my paper. The graph in question is here reproduced (Fig. 1); M_i stands for M_0 and N for P_s or $-P_s$, while the other notations are the same as in Schlitt's paper. As mentioned, the rigidity M_i/φ has a maximum, at $N\approx 2\pi^2EI/L^2$. Further, in accordance with Schlitt's statement, it is seen that the rigidity can be reduced to zero by a tensile force as well as by a compressive force. The latter force is however much larger than presumed by Schlitt, namely about 80 percent of the buckling load of the spring, which for clamped ends is equal to $4\pi^2EI/L^2$. The success of the actual application of this property will therefore be questionable.

On closer examination of the calculation given it is found that some signs in Eqs. (15), (16) and (17) are mistaken and that R_1 and R_2 must be replaced by $-R_1$ and $-R_2$ respectively. Further, the right-hand member of Eq. (18) must be put equal to $-\tan(L/\lambda)$ instead of to $\tan(L/\lambda)$. This explains the observed discrepancy.

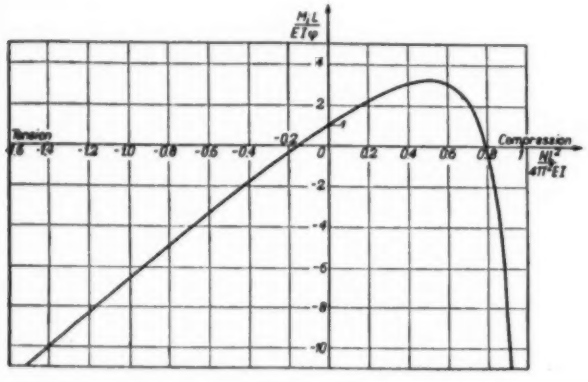

FIG. 1.

[1] H. Schlitt, "Investigations of a bearing for small angular deflections without any friction and with negligible restoring torque," J. App. Phys. 20, 297–301 (1949).
[2] J. A. Haringx, "The cross-spring pivot as a constructional element," Appl. Sci. Res. A1, 313–332 (1949).

Reply to J. A. Haringx

HELMUT SCHLITT
Fort Bliss, Texas
October 21, 1949

I AM very thankful to Mr. Haringx for calling my attention to a discrepancy between our investigations. Concerning the physical side of the problem, I entirely agree with his considerations. There must be a gradual transition between the torque curves plotted in Figs. 3 and 4 of my paper as far as they represent bearings of equal design. Since these curves refer to different design datas of a bearing (different value of R_1/R_2), I expected the maximum of the curve to be always on the opposite side of P_s/P_{s_0}. Actually there was a sign mistaken in my calculations of the compression bearing. Correcting this mistake my formulas show the changes stated by Haringx. The revised Eq. (18) yields then for the first solution for L/λ the value 5.59 corresponding exactly to the one taken from his graph.

This higher value for L/λ causes a correction of the design figures for the suggested bearing combination which is supposed to be independent of temperature. Nevertheless it seems possible to me to achieve the desired features. The fact that the compression force for zero torque of the compression bearing is rather close to the buckling load seems to me of minor importance as long as there are used three springs for each bearing. For such a design obviously no buckling can occur since the position of the two rings with respect to each other is invariably fixed.

Comments on Creep and Damping Properties of Polystyrene

WILLIAM N. FINDLEY
College of Engineering, University of Illinois, Urbana, Illinois
November 1, 1949

IN a recent paper,[1] Sauer, Marin, and Hsiao presented data on creep and damping capacity of polystyrene together with speculations on a correlation between damping and creep rate.

Interpretation of Creep Data: It may be of interest to consider another interpretation than the power equation used by the authors. An examination of the data given in the authors' Fig. 6[2] disclosed that the data might be represented about equally well by an exponential equation since the data formed nearly a straight line on a plot of log creep rate *vs.* stress. Because the exponential function is a good approximation of the hyperbolic sine function for relatively large values of the stress the latter function was suggested. The equation

$$v=0.26\sinh(\sigma/540) \qquad (1)$$

was found to represent the authors' data fairly well. This fact is of particular interest since there is a theoretical basis for expecting a hyperbolic sine relation which is to be found in studies of the activation process.[3]

In connection with these speculations it may be of interest to consider the creep rate data obtained by the writer[4] for tension creep tests at the same temperature and humidity used by the authors (see Fig. 1). These data were obtained from tests of a polystyrene made by the same manufacturer and having the same molecular weight of about 68,000 (Staudinger).

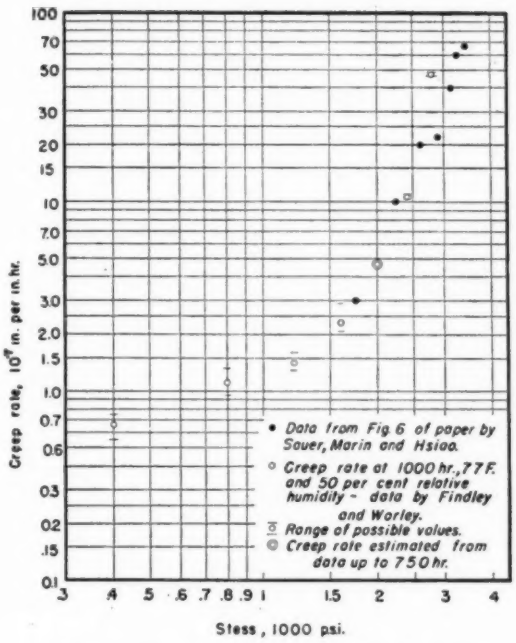
Creep rate, 10^{-6} in. per in. hr.
Stress, 1000 psi.
• Data from Fig. 6 of paper by Sauer, Marin and Hsiao.
○ Creep rate at 1000 hr., 77 F. and 50 per cent relative humidity - data by Findley and Worley.
Range of possible values.
Creep rate estimated from data up to 750 hr.

FIG. 1. Creep rate *vs.* stress for polystyrene in tension.

In Fig. 1 the data obtained by the authors as well as that obtained by the writer, are shown plotted on logarithmic coordinates. It is evident that the agreement is reasonably good over the interval of stress values covered in common by both tests. It is also apparent that the data obtained by the writer at the lower stresses indicate that the relationship between creep rate and stress was not a power function for the writer's tests, since the data form a curve rather than a straight line in Fig. 1.

The data obtained by the writer were represented more accurately by a hyperbolic sine function than a power equation. If the writer's datum at the largest stress is neglected, the remaining creep rates fit Eq. (1) within about $\pm 0.4\times10^{-7}$ in. per in. hr.

Interpretation of Dynamic Measurements: In view of the small scatter in the test data for damping presented by the authors, it seems appropriate to examine further the implications of these results as related to the test method and method of analysis employed, even though the method and analysis have been widely used to study the dynamic properties of many materials.

The equations used by these investigators were derived from the motion and forces present in resonant vibration of a single-degree-of-freedom system consisting of masses, linear spring, viscous damping and sinusoidal excitation. The differential equation for such a system is of the form[5]

$$M\ddot{x}+c\dot{x}+kx=F\sin\omega t, \tag{2}$$

where x, $\dot{x}$, and $\ddot{x}$ are the displacement, velocity, and acceleration respectively, M is the mass, c is the viscous damping coefficient, k is the stiffness of the spring, and $F\sin\omega t$ is the sinusoidal forcing function.

From the solution of Eq. (2) it is possible to calculate the energy dissipated per cycle, W_D, by the viscous damping.[6] The relationship obtained is

$$W_D=\pi c a^2\omega, \tag{3}$$

where a is the amplitude of vibration. For the mechanical system described above this dissipated energy is equal to the work done per cycle by the disturbing force F, as calculated by the authors, providing energy lost to the surroundings was negligible.

Equation (3) indicates that the energy dissipated per cycle by the damping is proportional to the square of the amplitude of the motion. If the stress, σ, in the specimen is understood to be defined in terms of the spring force alone and the spring force is linear as required in this solution, then the stress is proportional to the amplitude so that Eq. (3) becomes

$$W_D=C\sigma^2\omega, \tag{4}$$

where C includes the viscous damping constant and the dimensions and stiffness of the specimen.

That the tests reported by the authors showed the work done per cycle by the disturbing force (oscillator) to be proportional to the 2.3 power of stress may indicate that polystyrene behaves as a viscous material for damping. The fact that the 2.3 power was found instead of the 2.0 power may be a manifestation of the use of an equation for a system with a single viscous element, Eq. (2), to represent the behavior of a viscous material having a distribution of viscosities or relaxation times such as polystyrene is considered to have.

On the other hand it seems equally possible to the writer that the 2.3 power may result from some part of the mechanical equipment used by the authors not performing in accordance with the linear relationship required by the procedure used to calculate the damping capacity.

Further evidence that the test equipment may not function in the linear manner described by Eq. (2) are found in the facts that: (a) the so-called dynamic modulus calculated from the vibration test (5.3×10^5 p.s.i.) differs from the static modulus (4.5×10^5 p.s.i.) and in the opposite direction from that reported by previous investigators[7] using the same apparatus in tests of other plastics; (b) the dynamic modulus changed with stress. Neither of these conditions is in accord with Eq. (2) which requires the dynamic modulus to be constant and equal the static modulus.

Some of the possible sources of this disagreement may be: (A) the stiffness of the specimen changes with displacement; (B) the stiffness of the specimen changes with strain rate (or frequency); (C) the oscillator may not provide the required pure sine wave of disturbing force; (D) the damping in the specimen may not be viscous; and (E) there may be energy dissipated in non-linear friction in the grips or other connections. It is possible that the modulus increased *slightly* with increase in strain rate if the increase overlapped a portion of the spectrum of relaxation times for the material. But it seems more likely that the damping was not viscous.

An extensive study of the possible sources of extraneous energy dissipation in one type of damping testing machine has recently been reported.[8] These investigators found external losses of energy amounted to as much as 500 times the intrinsic energy dissipation of aluminum alloy specimens.

The following equation which has been considered for damping in structures[9] might prove to be more suitable in the present instance than the viscous equation for representing hysteresis damping:

$$M\ddot{x}+k(\cos b_1+i\sin b_1)x=Fe^{i\omega t}. \tag{5}$$

The damping in this equation, called complex damping, is represented in the equation by the phase angle of lag b_1.

A study of the solution of this equation shows that the damping capacity (energy absorbed per unit volume per cycle for a uniform rod vibration axially) is proportional to $\sin b_1$, the square of the stress, inversely proportional to the elastic modulus and independent of the frequency. The last three of these relationships are in accord with much of the reported data on damping capacity in materials: (A) Kimball and Lovell[10] found the damping capacity to be proportional to the square of the stress for a number of different materials. (B) They also gave rather conclusive evidence that the damping capacity is independent of frequency (or velocity). (C) The data on damping capacity reported in the literature indicates that in general materials with the highest elastic moduli have the lowest damping capacity and vice versa.

A comparison of the damping capacity described for complex damping with that predicted for viscous damping shows that the viscous damping capacity calculated from Eq. (3)

cannot be expressed (in terms of energy per unit volume) as a function of stress and factors which are properties of the unit volume of the material alone. Other quantities which depend on the particular configuration, such as the mass of the vibrating body and dimensions of the specimen tested, cannot be eliminated from the expression. Also the viscous damping capacity is not a function of the modulus of elasticity and is a function of frequency —the reverse of the relationship for complex damping.

Correlation of Damping Capacity and Creep Rate: It seems likely that damping and creep (also fatigue) are different responses to the same action at the atomic or molecular level of action in a material. Thus it would be likely that equations derived from the atomic or molecular behavior to describe these three different responses to mechanical loading would have at least one factor in common.

It would indeed be very useful if creep could be predicted from damping measurements. But to be dependable such predictions must be based on correspondence of fundamental behavior of the material under the two conditions.

The following observations may be of interest in trying to evaluate the correlation reported by the authors. For viscous behavior the damping capacity varies with the square of the stress and the creep rate varies with the first power of stress;[11] whereas the authors found a 2.3 power of stress for damping and 4.53 power of stress for creep rate.

The failure of the data for styrene to follow the same stress laws predicted for viscous behavior for both damping and creep does not seem too unreasonable in view of the differences in magnitude of strain and strain rate in the two phenomena. The behavior may result from quite different mechanisms, for example, rupture of secondary bonds between atoms in damping and movement during creep of groups of molecules of large molecular weight in a matrix of molecules of smaller molecular weight.

Dimensional considerations also suggest that the equations proposed by the authors are unlikely to be more than approximations of actual physical relations since a dimensional quantity raised to a variable or decimal power is dimensionally untenable.

[1] Sauer, Marin, and Hsiao, J. App. Phys. **20**, 507 (1949).
[2] The plotted values for two points seem to differ somewhat from the values in their Table II.
[3] Dushman, Dunbar, and Huthsteiner, "Creep of metals," J. App. Phys. **15**, 108–124 (1944).
[4] W. N. Findley and W. J. Worley, "Creep and fatigue of polystyrene as influenced by temperature" (not yet published).
[5] Timoshenko, *Vibration Problems in Engineering* (D. Van Nostrand Company, Inc., New York, 1937), second edition, p. 38.
[6] See reference 5, p. 45.
[7] B. J. Lazan, "Some mechanical properties of plastics and metals under sustained vibrations," Trans. Am. Soc. Mech. Eng. **65**, 87 (1943). J. M. Robertson and A. Yorgiadis, "Internal friction in engineering materials," J. App. Mech. **68**, A173 (1946).
[8] Cottell, Entwistle, and Thompson, "The measurement of the damping capacity of metals in torsional vibration," J. Inst. Metals, March 1948, p. 373.
[9] N. O. Myklestad, *Vibration Analysis* (McGraw-Hill Book Company, Inc., New York), second edition (in preparation), Section IV-5.
[10] A. L. Kimball and D. E. Lovell, "Internal friction in solids," Phys. Rev. **30**, 947–959 (1927).
[11] Schmidt and Marlies, *Principles of High-Polymer Theory and Practice* (McGraw-Hill Book Company, Inc., New York, 1948), Chapter 7.

On the Growth of Metallic Crystals from the Vapor Phase

GEORGE R. KEEPIN, JR.*
Massachusetts Institute of Technology, Cambridge, Massachusetts
November 7, 1949

GENERALLY speaking there are four methods of growing crystals: (1) from the melt; (2) from solution; (3) from metastable crystals in contact with the growing crystal in a continuous solid; and (4) from the vapor. Only the first three of these methods have been studied extensively. Recently, however, there has been increasing interest in the possibilities of the last method for growing large single crystals from certain metals and metal salts. Parameters affecting the growth of metallic crystals from their vapor phase are being studied in the hope that the results may indicate generally the conditions for growth of other metallic and metallic-halide crystals by this comparatively new technique.

TABLE I.

	Temperature of vapor gun source	Partial pressure of vapor source	Temperature of collecting surface	Pressure of nitrogen atmosphere
Cadmium	310–321°C	$\sim 10^{-1}$ mm	305±3°C	$10^{-2}-5\times10^{-2}$ mm
Zinc	409–419°C	$\sim 2\times10^{-1}$ mm	410±3°C	$10^{-2}-3\times10^{-2}$ mm

All crystallization processes may be subdivided into two main factors: the velocity of appearance of nuclei and the subsequent growth of these nuclei. The experimental conditions which have greatest effect on these two factors have been given by Sakui[1] as (1) density of vapor or temperature of vapor source, (2) temperature of collecting surface, and (3) pressure of atmosphere in which crystal growth occurs. These effects have been investigated in the case of metallic cadmium and zinc crystals grown from their vapor.

An internally heated molybdenum crucible containing the metal charge was used as a gun-type vapor source. Temperatures were measured at various points in the molybdenum gun and collector plate by copper-constantin thermocouples. The entire assembly was contained within an evacuated ten-inch bell jar. All work was done in an atmosphere of oil-pumped (dry) nitrogen which was freed of oxygen by passage through fine copper mesh at 375°C. In order to obtain the required degree of purity in the charge it was found necessary to use the spectroscopically pure metal which was then further purified by a multiple distillation procedure similar to that used by Overhage[2] in the purification of sodium metal. A silver collecting surface was found most satisfactory for cadmium while a copper or nickel surface proved best for the zinc runs. For convenience in taking data, observed crystalline forms were divided into eight categories ranging from a dense amorphous-like deposit to well-developed single crystals. Vapor gun temperatures were varied in five degree intervals from 30° below to 30° above the melting point of the metal under study. Nitrogen pressure was varied in appropriate steps from 10^{-5} to 760 mm. Collector plate temperature was increased in five degree intervals from room temperature to 10° above the melting point of each metal.

A predominance of dense amorphous-like deposit was found at low collector plate temperatures. This is to be expected since a metal vapor will condense (form growth nuclei) fastest where the difference between the vapor pressure of the solid phase and the supersaturated vapor is greatest. X-ray powder patterns of these amorphous-like deposits showed them to be microcrystals of mean diameter the order of 100A. Such studies indicated that no true amorphous deposits of either Cd or Zn were obtained in any of the work.

Conditions for the further growth of nuclei once they are formed are entirely different from those conditions necessary for their formation; thus the most perfectly developed single crystals were obtained at collector plate temperatures (T_{cp}) slightly below the respective melting points, *viz.*, optimum T_{cp} for Cd is 305±3°C and for Zn is 410±3°C. As T_{cp} is decreased from these optimum values, the amount of cadmium decreases steadily while the amount of zinc deposit, predominantly crystalline powder, was actually found to increase reaching a maximum at $T_{cp}=350$°C; thus we may say that the smaller zinc atoms possess a greater nuclei-forming ability than cadmium atoms in this temperature range. As T_{cp} is raised above its optimum value a rapidly diminishing yield of successively smaller crystals is obtained so that at five degrees above optimum T_{cp} values only traces of microcrystals are obtained in the case of either metal. These temperature effects are quite explicable: the mobility of atoms adsorbed by the growing crystal surface determines their rate of deposition in the crystal lattice. Accordingly, high T_{cp} means high mobility and high

rate of orientation of atoms within the lattice matrix. The potential energy of properly oriented atoms is a minimum and their chance of further migration is vanishingly small so they are effectively removed from the adsorption layer and may be replaced by new atoms arriving from the vapor gun. Orderly single crystal growth requires a dynamic balance between these two processes maintaining thereby a constant density of atoms in the adsorption layer.

For nitrogen pressures greater than 0.5 mm only traces of crystalline powder are obtained. As the pressure is decreased from 0.5 mm larger but imperfect crystalline deposits result until in the pressure interval 10^{-2} to 10^{-1} mm dendritic crystals are predominant at intermediate T_{ep} (250–350° for Cd and 200–350° for Zn). Finally, at optimum T_{ep} values well-developed hexagonal crystals were obtained in the pressure interval 10^{-2} to 5×10^{-2} mm for Cd and 10^{-2} to 3×10^{-2} for Zn. At lower pressures only small yields of dendritic, inhomogeneous crystals were found. These pressure effects become quite plausible when we recall that the surface temperature of a growing crystal is higher than that of the crystal interior due to heat of solidification and that the growth of the crystal is largely controlled by the rate of heat transfer from the growing surface. Since in a given volume cooling by convection depends critically on pressure we may expect to obtain the desired thermal equilibrium within the adsorption layer at some pressure which we here call optimum, i.e. on the crystal surface the rate of heat increase by solidification equals the rate of heat loss principally by convection.

In summary, optimum conditions have been found for the growth of large, single crystals of cadmium and zinc as shown in Table I.

Electromagnetic vibration of the collector plate at a proper frequency may represent a distinct improvement over the present method of employing thermal (random) agitation to aid atomic orientation in the growing crystal lattice. The study of other vapor crystallization processes using such a technique is contemplated.

The author is indebted to Professor D. C. Stockbarger for suggesting this problem. This research was supported in part by the Harshaw Chemical Company, Cleveland, Ohio.

* Now at Radiation Laboratory, University of California, Berkeley.
[1] S. Sakui, Tokyo Inst. Phys. Res. **34**, 1131 (1938).
[2] Overhage, *Procedures in Experimental Physics* (Prentice-Hall, Inc., New York, 1945), p. 534.

Thermionic Emitting Properties of Thoria-Rhenium*

G. A. Espersen
Philips Laboratories, Inc., Irvington-on-Hudson, New York
December 12, 1949

SINTERED conducting type thoria emitters containing tungsten have been investigated by Fan.[1] Similar thoria emitters containing 250 mesh pure rhenium powder furnished by the University of Tennessee, Knoxville, Tennessee, have been studied by the writer.

Emitters consisting of 50 percent thoria—50 percent rhenium were fabricated into rods by extrusion, after which they were fired successively at 1500° C_b for 15 minutes in wet hydrogen and 2000° C_b for 15 minutes in dry hydrogen.

The emitter rods were assembled into guard ring type diodes using molybdenum anodes.

Thermionic emission data taken on a number of diodes indicated average values of $A=11.7$ amp/cm²/deg. K², $\phi=2.83$ ev for the Richardson equation.

It is interesting to note that the thermionic emission from thoria-rhenium at 1630°K is approximately 55 milliamperes while thoria-tungsten yields 0.2 milliampere/cm² at the same temperature. At 2040°K the emission for thoria-rhenium is 1.2 amp/cm² while thoria-tungsten is 0.88 amp/cm².

Activation of both the thoria-rhenium and thoria-tungsten emitters is effected by flashing them at 2400°K for approximately 30 seconds.

After 150 hours of life test at 2040°K the thoria-rhenium diodes indicated a smaller deposit of vaporized materials upon the walls of the glass bulb then the thoria-tungsten diodes.

* Partial fulfillment of ONR Contract No. N6onr-261, Task Order 1.
[1] H. Y. Fan, J. App. Phys. **20**, 682 (1949).

Activation Energy for Copper Self-Diffusion

John H. Hollomon
General Electric Research Laboratories, Schenectady, New York
November 28, 1949

I HAVE read the paper by Dedrick and Gerds in the November, 1949, *Journal of Applied Physics*. From their data, plotted in their Fig. 5, they calculate 55,000 calories per gram mole—a heat of activation for self-diffusion in copper by taking the slope of the straight line they have drawn in that figure. A straight line through the points gives a heat of activation of approximately 40,000 calories per gram mole. This value is significantly smaller than any other previous experimental determinations obtained by widely accepted techniques.

Reply to "Activation Energy for Copper Self-Diffusion"

J. H. Dedrick and A. Gerds
Sylvania Electric Products, Bayside, Long Island, New York
December 19, 1949

THE authors appreciate the comment of Dr. J. H. Hollomon regarding the heat of activation for the self-diffusion of copper. It is true that if a line is drawn through all of the points in Fig. 5, the heat of activation is about 40,000 calories per gram mole as contrasted to the value of 55,000 calories per gram mole given in the paper. A recalculation by the authors shows the value to be 43,000 calories per gram mole. This is low, it is agreed, but close to the value of 46,800 calories per gram mole obtained by Raynor, Thomassen, and Rouse.[1]

However, the reason for drawing the line as shown in Fig. 5 was twofold. First, the interface radii obtained at 700°C were the smallest and, therefore, the least accurate, from the standpoint of measurement. Second, as Kuczynski[2] has pointed out, surface diffusion tends to become more effective at lower temperatures, which would effectively raise D_v. Therefore, the 700°C point was not weighted as heavily as the higher temperature points. The line as drawn thus represents the higher temperature data most rigidly.

[1] Raynor, Thomassen, and Rouse, Trans. Am. Soc. Metals **30**, 313 (1942).
[2] G. C. Kuczynski, J. Metals, 169 (February, 1949).

Study of Crack Propagation Using High Speed Motion Pictures

Herbert I. Fusfeld and Josephine Carr Feder
Pitman-Dunn Laboratory, Frankford Arsenal, Philadelphia, Pennsylvania
November 30, 1949

HIGH speed motion pictures have been used to observe the process of deformation and rupture in metals at high strain rates. Preliminary studies of fracture were conducted by utilizing flat tensile specimens ½ in. wide and $\frac{1}{32}$ in. thick with grids photographically imprinted on a sensitized surface. A fixture to hold the specimens was placed in a crank press used in drawing metal sheet, and operated at a speed of 960 in./min. A Fastex camera

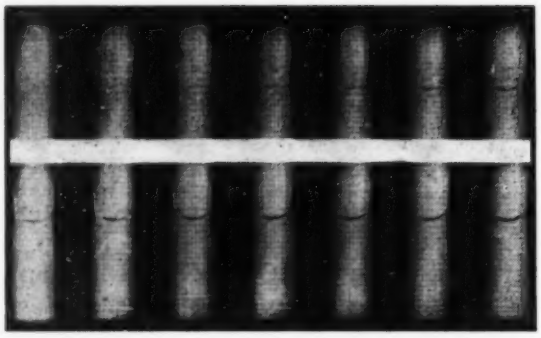

FIG. 1. Propagation of crack in aluminum strip under tension. Aluminum 75 SO, flat, 0.5 ×0.031 in. Time between frames = 102 μsec.

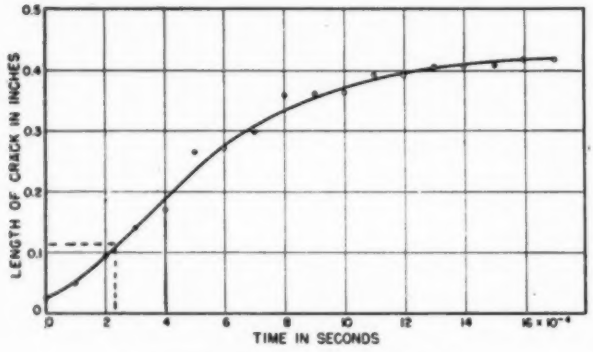

FIG. 2. Crack propagation as function of time.

was used to take motion pictures at a rate of 10,000 frames per sec.

Most of the aluminum and brass specimens tested broke along a line inclined at approximately 60° to the axis of the specimen. This break occurred too rapidly to be observed by the camera, i.e., in less than 10^{-4} sec. Thinning of the specimen along the line of fracture should be observed for approximately 5×10^{-4} sec. preceding the break.

Interesting observations have been made on the few specimens, three out of seventeen, which fractured along an irregular curve. The major portion of one such fracture is shown in Fig. 1 for aluminum 24 *ST*, showing a crack starting at the center and propagating to the edges. No noticeable thinning occurred prior to the appearance of the crack, indicating a more brittle behavior of the material than in the case where a straight line break occurs.

Measurements of the propagation of the crack are plotted in Fig. 2. The most interesting portion of the curve is at the start of propagation, as one can assume that the crack is in a uniform stress field and lines in the direction transverse to the axis of

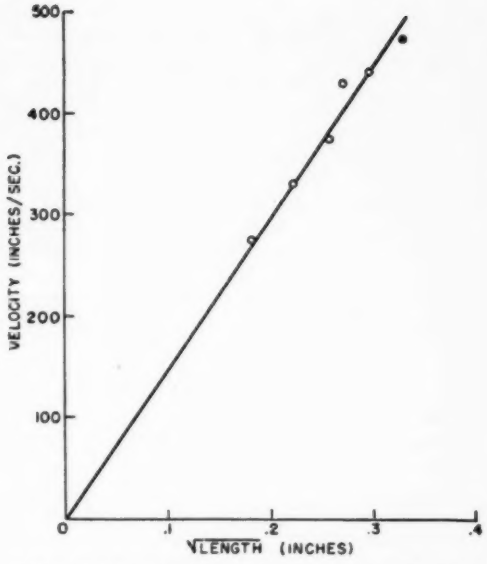

FIG. 3. Velocity of crack propagation *vs.* square root of crack length.

stress. Under these conditions, one can be guided by elasticity theory as applied to a narrow crack[1] to expect that the stress at the ends along the crack axis will increase as the crack length increases. It is therefore to be expected that the velocity of propagation would increase with the crack length, some data on which has appeared in the literature.[2]

To observe the dependence of velocity on crack length during the initial stages of propagation, the slope of the curve in Fig. 2 was measured at six points at which the ratio of crack length to specimen diameter varied from 0 to 0.64. The portion of the curve used is indicated on Fig. 2 within the dotted lines and the coordinate axes. Beyond this portion, the crack begins to extend in a direction away from that normal to the specimen axis, and the significance of "crack length" is uncertain.

A plot of velocity *vs.* length of crack on log-log paper gave a straight line with a slope of 0.47 computed on the principle of least squares. The same data is plotted in Fig. 3 as velocity *vs.* square root of crack length. A straight line through the origin provides a good fit. The slope computed by least squares is 1432 ± 26 in.$^{\frac{1}{2}}$/sec. It is therefore possible to represent the initial stages of crack propagation for the specimen of Fig. 1 by the expression:

$$v = k(s)^{\frac{1}{2}},$$

where v = velocity in in./sec.; k = slope in in.$^{\frac{1}{2}}$/sec.; s = crack length.

[1] S. Timoshenko, *Theory of Elasticity* (McGraw-Hill Book Company, Inc., New York), p. 177.
[2] G. Irwin, "Fracture dynamics," ASM Symposium on Fracturing of Metals (1948), p. 159.